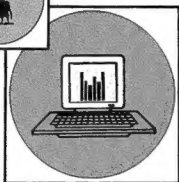
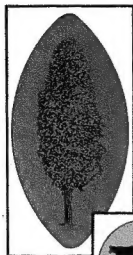


# الإحصاء في البحوث العلمية



محمد أبو يوسف

أستاذ الإحصاء بقسم الرياضيات  
كلية التربية جامعة عين شمس

الناشر



المكتبة الأكاديمية



# الإحصاء

## في البحوث العلمية





# الإحصاء في البحوث العلمية

محمد أبو يوسف

استاذ الإحصاء بكلية التربية  
جامعة عين شمس

الناشر



المكتبة الأكاديمية

### حقوق الطبع والنشر محفوظة

لا يجوز استنساخ أى جزء من هذا الكتاب أو إيجزائه بأى وسيلة إلا بإذن خطى من الناشر  
١٢١ ش التحرير — الدق — القاهرة ت : ٣٤٩١٨٩٠ / ٣٤٨٥٢٨٢ نلكس : ٩٤١٢٤ ABCMN UN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

قُلْ لَقَدْ أُوحِيَ إِلَيَّ وَوُحِّىَ إِلَيَّ وَوُحِّىَ إِلَيَّ  
فَلْيَأْمُرِ الْعَالَمِينَ

مَنْعَهُ اللَّهُ الْعَظِيمُ



# بسم الله الرحمن الرحيم

## مقدمة

يتناول هذا الكتاب أسس التحليل الإحصائي للتجارب الميدانية والمعملية في مجال العلوم . ولما كانت الدراسة المثمرة للإحصاء التطبيقي تتطلب حدا أدنى من المعرفة بالنظرية الإحصائية ، وإغفال ذلك يؤدي إلى فهم سطحي تنجم عنه أخطاء جسيمة في التطبيق العملي ، فقد عنى الكتاب بإرساء أساسيات وركائز هذه النظرية كما عنى بالربط بين النظرية والتطبيق وتقديم الأصول العلمية لشروط ومحددات ما يستخدم من طرق واختبارات وعمليات استدلال . وقد تجنب في ذلك كله الدخول في التفاصيل والبراهين الرياضية التي قد تعوق القارئ عن متابعة مسيرة التسلسل المنطقي للمفاهيم والأساليب الرئيسية المنشودة .

وقراءة هذا الكتاب لا تتطلب إلا اليسير من الخلفية الرياضية التي لا تخرج عن المبادئ الحسابية والجبرية البسيطة . وفي الموضوعات التي تحتاج إلى جهد كبير في إجراء العمليات الحسابية اللازمة للتحليل اعتمد الكتاب على الحاسب الالكتروني توفيراً للجهد وضماناً للدقة والسرعة ، دون أن يستلزم ذلك أن يكون الباحث قادراً على تشغيل الحاسب بنفسه بل يكفيه الاتصال بأحد مراكز الحاسب التي أصبحت اليوم متوافرة في الجامعات ومراكز البحث العلمي .

ولابد من الإشارة هنا إلى أن هذا الكتاب هو في حقيقة أمره طبعة مزيّدة من كتاب سبق لي تأليفه بعنوان « مقدمة في الإحصاء البيولوجي » تكفلت جامعة

قطر مشكورة بإصداره سنة ١٩٨٤ مساهمة منها في المد الثقافي العربى وتبادل الكتب والمطبوعات العلمية مع الجامعات العربية والهيئات الثقافية الأخرى . ولقد سعدت بتدريس ذلك الكتاب عدة سنوات خلال فترة عملى بكلية العلوم بتلك الجامعة لطلاب من مختلف الشعب العلمية فى مرحلة البكالوريوس وفى الدراسات العليا التمهيدية . ولقد ضاعفت الزيادة فى هذا الكتاب من حجم الكتاب السابق ، وهى تتمثل فى توسيع بعض الفصول خاصة فصلى تحليل التباين والانحدار الخطى البسيط ، وكذلك فى إضافة خمسة فصول جديدة تحمل طابعا متقدما هى الفصول السابع والحادى عشر والثانى عشر والثالث عشر والخامس عشر . وتستهدف هذه الزيادة استكمال الجوانب التطبيقية لعلم الإحصاء تعزيزا لمحتوى الكتاب بما يخدم احتياجات قطاع أكبر من الطلاب والباحثين ، ومساهمة فى سد إحدى الثغرات العديدة التى تعاني منها المكتبة العربية فى مجال العلوم .

ويرجو المؤلف أن يكون فى الأسلوب الذى قدم به المادة وما استخدمه من أمثلة مستمد أغلبها من بحوث ودراسات واقعية ما يمكن القارئ من اكتساب منهج فكرى فى التحليل الإحصائى يجعله قادرا على المساهمة فى تخطيط التجارب وتحليلها ، والاعتماد على نفسه فى الاستزادة من دراسة الإحصاء التطبيقى ، وفى تفهم ما ينشر من البحوث فى هذا الميدان . ولعل فى تقديم المصطلحات الإحصائية باللغتين العربية والانجليزية ما ييسر للقارئ متابعة المراجع والبحوث الأجنبية التى لا غنى للباحث عنها .

أسأل الله التوفيق وعلى الله قصد السبيل .

المؤلف

محمد أبو يوسف

١٩٨٩/٨/١٨

## المحتويات

### الصفحات

٣٩ - ١٥	الفصل الأول : مفاهيم أولية .....
	المجتمع والعينة - العينة العشوائية البسيطة - المتغيرات
	الإحصائية - الأخطاء العشوائية - قواعد تقريب الأعداد -
	الاحتمال - النماذج الرياضية .
٨٥ - ٤١	الفصل الثاني : التوزيعات التكرارية .....
	الجداول التكرارية - التمثيل البياني للتوزيعات التكرارية - التمثيلات
	والربيعات - الوصف العددي للتوزيعات التكرارية ( الوسط
	الحسابي والانحراف المعياري ، الالتواء ، التفرطح ) - شكل
	الساق والورقة .
١٢١ - ٨٧	الفصل الثالث : بعض نماذج الاحتمال .....
	توزيعات الاحتمال الثابتة - نماذج الاحتمال - توزيع ذي الحدين -
	تقدير الدليل 2 - توفيق توزيع ذي الحدين لتوزيع تكراري
	معلوم - توزيع بواسون ( كتقريب لتوزيع ذي الحدين -
	كنموذج لتوزيع الأحداث النادرة ) - اختبار استقلال الأحداث
	النادرة - غمط التجمع وغمط التناثر - توزيع باسكال -
	التوزيع الهندسي - توزيعات الاحتمال المتصلة .
١٤٧ - ١٢٣	الفصل الرابع : التوزيع المعتدل والتوزيع اللوغاريتمي .....
	( أولا ) : التوزيع المعتدل - بعض خواص التوزيع - جداول

المساحات أسفل المنحنى المعتدل المعيارى - الكشف عن  
الاعتدالية - طريقة بيانية للكشف عن الاعتدالية - معالجة  
عدم اعتدالية التوزيع - تقريب توزيع ذى الحدين بتوزيع  
معتدل .

( ثانيا ) : التوزيع المعتدل اللوغارىتمى - بعض خواصه .

الفصل الخامس : توزيعات خاصة ..... ١٤٩ - ١٥٨

توزيعات ت ،  $\chi^2$  ، ف وجداول القيم الحرجة .

الفصل السادس : نظرية العينات ..... ١٥٩ - ٢٢٦

نظرية العينات - توزيعات المعاينة - الإحصاءة - الخطأ المعيارى -  
التقدير غير المتحيز - المعاينة من مجتمعات معتدلة - المعاينة  
من توزيع ذى الحدين - اختبارات الفروض - اختبارات  
( اختبار فرض عن الوسط الحسابى لمجتمع معتدل - فترات الثقة  
للووسط الحسابى لمجتمع معتدل - مقارنة متوسطى مجتمعين  
معتدلين - اختبار استقلال الأحداث النادرة ) - اختبار  $\chi^2$   
( اختبار فرض عن توزيع مجتمع - اختبار حسن المطابقة -  
اختبار استقلال خاصيتين - فترات الثقة لتباين مجتمع معتدل ) -  
اختبار ف - فترات الثقة للنسبة في مجتمع والفرق بين نسبتي -  
تحديد حجم العينة - مراقبة الإنتاج .

الفصل السابع : حساسية اختبارات الفروض ..... ٢٢٧ - ٢٥٨

نوعا الأخطاء الإحصائية - حساب قيمة  $\beta$  في اختبار ذى جانب  
واحد عن متوسط مجتمع معتدل ( دالة القوة - العوامل المؤثرة  
في الخطأ من النوع الثانى - كيفية زيادة قوة الاختبار ) -  
حساب قيمة  $\beta$  في اختبار ذى جانبيين عن متوسط مجتمع  
معتدل - حساب قيمة  $\beta$  في اختبار الفرق بين مجتمعين  
معتدلين - حساب قيمة  $\beta$  في اختبار النسبة .



## الفصل الثامن : تحليل التباين وتصميم التجارب ..... ٢٥٩ - ٣٦١

التحليل الإحصائي وتصميم التجارب - تحليل التباين - مصطلحات وتعريف - التجارب ذوات العامل الواحد - المقارنة القبلية والبعدية بين المتوسطات - التجارب ذوات العاملين المتفاعلين وغير المتفاعلين - المقارنات التزاوجية - التجارب ذوات الثلاثة العوامل ( في مربع لاتيني ) - معالجة الانحرافات عن افتراضات التحليل - عودة إلى مقارنة المتوسطات - المقارنات القبلية - معيار استقلال مقارنتين - اختبار المقارنات القبلية - تجزئء مجموع المربعات - اختبار المقارنات البعدية - النموذج عشوائى التأثيرات .

## الفصل التاسع : الانحدار الخطى البسيط ..... ٣٦٣ - ٤١٢

المجتمع ذو المتغيرين - شكل الانتشار - تقدير البارامترين  $\alpha$  ،  $\beta$  طريقة المربعات الصغرى - الخطأ المعياري لخط الانحدار - استنتاجات إحصائية - التوسع في استخدام الانحدار الخطى البسيط - معنى آخر للانحدار - الاختلاف المفسر وغير المفسر - معامل التحديد - تحليل الانحدار حين وجود أكثر من قيمة ص لكل قيمة س - اختبار جودة العلاقة الخطية - ملاحظات عن افتراضات الانحدار - استخدامات الانحدار .

## الفصل العاشر : الارتباط الخطى البسيط ..... ٤١٣ - ٤٣٩

الانحدار والارتباط - افتراضات الارتباط الخطى البسيط - معامل الارتباط العزى ( بيرسون ) - مميزات معامل الارتباط العزى - دلالة معامل الارتباط العزى - اختبار  $r = 0$  - اختبار  $r = 1$  - اختبار دلالة الفرق بين معاملى ارتباط عيتين مستقلتين - التمييز بين الانحدار والارتباط في دراسة

- المشكلات - معامل ارتباط الرتب ( سيزمان ) - مميزات  
معامل ارتباط الرتب - دلالة معامل ارتباط الرتب .
- الفصل الحادى عشر : تحليل التباين ..... ٤٤١ - ٤٦٨
- التباين - العلاقة بين تحليل التباين وتحليل التباين - النموذج  
الإحصائى - خطوات تحليل التباين - المقارنة بين المتوسطات  
المعدلة - تحليل التباين فى التجارب ذوات العاملين - المقارنة  
بين أزواج المتوسطات .
- الفصل الثانى عشر : الانحدار والارتباط الخطى المتعدد ..... ٤٦٩ - ٥١٠
- ( أولا ) الانحدار الخطى المتعدد - كامتداد للانحدار الخطى  
البسيط - إيجاد معادلة الانحدار - إيجاد الخطأ المعيارى للتقدير  
من خط الانحدار - اختبار دلالة الانحدار ككل - اختبار دلالة  
معاملات الانحدار الجزئية - استخدام الحاسب الالكترونى -  
طريقة دوليتل لحل المعادلات الخطية - أسلوب آخر لاختبار  
دلالة معاملات الانحدار الجزئية - اختبار متغيرات التنبؤ .
- ( ثانيا ) الارتباط الخطى المتعدد : معامل الارتباط الخطى المتعدد  
واختبار دلالة - معامل الارتباط الجزئى واختبار دلالة .
- الفصل الثالث عشر : دالة التمييز ..... ٥١١ - ٥٢٨
- دالة التمييز - إيجاد دالة التمييز ( حالة متغير واحد وحالة ك من  
المتغيرات ) - افتراضات التحليل - النقطة الحدية - احتمال خطأ  
التقسيم - اختبار تساوى أزواج المتوسطات - اختبار ت<sup>٢</sup> -  
استخدامات دالة التمييز .
- الفصل الرابع عشر : الطرق غير البارامترية ..... ٥٢٩ - ٥٦٣
- اختبار التلاحقات - اختبار الإشارة - اختبار ويلكوكسن  
للمقارنات الزوجية - اختبار الاتجاه - اختبار كرومكال -  
واليس ، اختبار فريدمان

الفصل الخامس عشر : اختيار العينات وتحليلها ..... ٥٦٥ - ٦٠٦

المعينة العشوائية - المعينة الاحتمالية - العينة العشوائية البسيطة -  
العينة الطبقية ( طريقة التقسيم المتناسب - طريقة التقسيم  
الأمثل - تقدير البارامترات والأخطاء المعيارية - المعينة  
الطبقية من مجتمع ذى حدين ) - العينة المتعددة المراحل  
( التحليل الاحصائي - الاختبارات الإحصائية ) - العينة  
المنتظمة - المعينة المساحية - العينات غير الاحتمالية .

الملحق (١) أجوبة التمارين ..... ٦٠٧ - ٦١٧

الملحق (٢) جداول إحصائية ..... ٦١٩ - ٦٤١

مراجع . ..... ٦٤٣



## الفصل الأول

### مفاهيم أولية

يتناول هذا الفصل عدداً من المفاهيم التي تركز عليها دراسة الإحصاء التطبيقي خاصة في مجال العلوم ، فهي بمثابة الأبنية التي لا مفر من إرسائها قبل متابعة تلك الدراسة . وهي وإن بدت منفصلة عن بعضها إلا أنها تتكاتف معا لخدمة موضوعات الفصول التالية .

#### ( ١ - ١ ) المجتمع والعينة :

إن التمييز بين المجتمع والعينة هو أول ما ينبغي أن يتنبه له أي باحث تطبيقي خاصة حين يستخدم الطرق الإحصائية وعملية الاستدلال الإحصائي .

في الإحصاء تستخدم كلمة مجتمع للتعبير عن أي مجموعة ( منتهية أو لا نهائية ) من الأشياء أو الأحداث التي تكون موضع اهتمامنا في وقت ما من حيث ظاهرة ما أو متغير ما سـ . ومن أمثلة المجتمعات الإحصائية : جميع نباتات حنك السبع المزروعة في حقل ما في وقت ما ، جميع نباتات البسلة في العالم ، جميع القواقع في بحيرة ما ، جميع الفئران من نوع معين ، سقوط المطر في منطقة ما ...

وينبغي أن يكون المجتمع الذي ندرسه معرّفاً تعريفاً جيداً خاصة فيما يتعلق بالمتغير سـ وطريقة قياسه وفي تحديد الوحدة أو العنصر التي يتكون المجتمع من مجموعها . فمثلاً قد يعبر المتغير سـ عن الطول أو الوزن أو اللون أو عدد ضربات القلب التي تنتج عن حقن قلب الفأر بمادة الأدرينالين ... وقد تكون الوحدة هي قرن بسلة أو قوقعة أو قلب فأر ...

وفي كثير من الأحيان يمكن وصف المتغير  $x$  في مجتمع ما عن طريق نموذج نظري يوضع على هيئة معادلة أو صيغة رياضية تعبر عما يسمى « توزيع المجتمع » فنقول مثلا إن مجتمعنا ما هو مجتمع معتدل أو مجتمع ذو حدين أو مجتمع بواسوني ... وهذا ما سوف نتناوله فيما بعد .

والأعداد التي تميز مجتمعنا ما تسمى بارامترات ( أو معالم أو أدلة أو ثوابت ) المجتمع وهي أعداد ثابتة تميز كل مجتمع عن غيره من المجتمعات حتى ولو كان له نفس التوزيع ، مثل الوسط الحسابي  $\mu$  وهو مقياس للنزعة المركزية للمجتمع ، والانحراف المعياري  $\sigma$  وهو مقياس لتشتت مفردات المجتمع حول الوسط الحسابي .

أما العينة فهي جزء من المجتمع يختار بحسب مواصفات معينة وبهدف استخدامها للدراسة المجتمع ، وهناك من النظريات والطرق الإحصائية ما يمكننا من تقدير بارامترات المجتمعات أو مقارنتها أو إصدار قرارات أو تنبؤات بشأنها عن طريق فحص ودراسة عينات مأخوذة منها .

وبطبيعة الحال ينبغي أن تختار العينة بحيث تمثل المجتمع أفضل تمثيل ممكن ، على أن التحليل الإحصائي يتطلب بالضرورة أن تكون العينة عشوائية . والعشوائية لا تعني أن نأخذ جزءا من المجتمع بشكل جزائي ، بل هي إجراء يصمم بدقة بحيث يضمن عدم وجود تحيز من أى نوع قد يؤثر في عملية اختيار العينة وبحيث يكون لكل وحدة من وحدات المجتمع احتمال معروف للدخول في العينة . ومن ثم فقد وضعت خطط مختلفة للمعاينة العشوائية random sampling plans أى لسحب العينات العشوائية ، يتوقف استخدام أى منها لدراسة مجتمع ما على طبيعة هذا المجتمع وعلى الهدف من دراسته . ومن أشهر هذه الخطط ما ينتج العينات التي تحمل الأسماء الآتية : العينة العشوائية البسيطة - العينة الطبقية - العينة ذات المراحل - العينة المنتظمة - العينة المساحية ... وسنهم هنا بصفة خاصة بالعينة العشوائية البسيطة التي هي على أية حال أساس لكثير من تلك الخطط ، أما بقية الخطط فقد أفرد لها فصل خاص هو الفصل الخامس عشر من هذا الكتاب .

## ( ١ - ٢ ) العينة العشوائية البسيطة : SIMPLE RANDOM SAMPLE

لعل أبسط تعريف للعينة العشوائية البسيطة هو أنها تلك العينة التي تؤخذ من المجتمع بحيث يكون لكل وحدة من وحداته نفس الفرصة في الظهور في العينة . ولذلك فإن هذه العينة لا تصلح لتمثيل المجتمع إلا إذا كان هذا المجتمع متجانسا من حيث المتغير الذي ندرسه . وفيما يلي نذكر كلمة عينة فسوف نقصد عينة عشوائية بسيطة .

ولعل أوضح خطة لاختيار عينة عشوائية بسيطة من مجتمع منتهي هي تلك التي اشتهر استخدامها في تحديد أصحاب جوائز المسابقات التليفزيونية . نفرض مثلا أن لدينا ٤٥٢ شخصا أجابوا إجابات صحيحة ونريد أن نختار منهم ٢٠ شخصا دون تحيز . إن هذا يعني أن لدينا مجتمعا متجانسا حجمه ٤٥٢ ونريد سحب عينة عشوائية بسيطة حجمها ٢٠ . نعد ٤٥٢ قطعة صغيرة متطابقة من الورق ونكتب عليها أسماء هؤلاء الأشخاص ( أو نعطيهم أرقاما متسلسلة من ١ إلى ٤٥٢ ) ثم نكتب هذه الأرقام على قطع الورق . نضع هذه القطع في علبة ونخلطها جيدا ثم نسحب منها ٢٠ قطعة الواحدة بعد الأخرى دون أى تحيز مع إرجاع كل قطعة تسحب إلى الصندوق قبل كل سحبة ونخلط القطع جيدا في كل مرة وإعمال القطع التي تتكرر في السحب ، فتكون مجموعة الأشخاص الذين تظهر أرقامهم أو أرقامهم على هذه القطع هي العينة المطلوبة .

غير أنه يمكن الاستغناء عن العلبة وقطع الورق وبناء خطة الاختيار على أحد الجداول المسماة بمجداول الأرقام العشوائية ، مثل الجدول (١) بملحق هذا الكتاب . ويتألف هذا الجدول من عدة أعمدة ( أو صفوف ) مجموعة كل حمدة معا للسهولة وبكل عمود أرقام ٠ ، ١ ، ٢ ، ... ، ٩ مرتبة ترتيبا خاصا يجعل لكل من هذه الأرقام نفس الفرصة في الظهور في أي موضع بالجدول ، ويمكن أن نقرأ منها أعدادا متتابعة يتألف كل منها من رقم واحد أو رقمين أو ثلاثة .. أى نأخذ عمودا واحدا أو اثنين أو ثلاثة ... حسبما نريد ، كما يمكن أن نبدأ القراءة

من أى صف أو أى عمود وفي أى اتجاه ، إلى أعلى أو إلى أسفل يمينا أو يسارا أو قطريا . ويعتمد عدد الأعمدة التي نختارها على عدد الأرقام التي يشتمل عليها حجم المجتمع .

### مثال ( ١ - ١ ) :

لدينا مجتمع حجمه ٥٠٠ من المرضى بمرض معين ونريد أن نستخدم جدول الأرقام العشوائية لاختيار عينة عشوائية بسيطة حجمها ٢٠ لإجراء بعض القياسات على عناصرها .

نظرا لأن حجم المجتمع وهو ٥٠٠ يتألف من ثلاثة أرقام ، نعطي لوحات المجتمع أرقاما مسلسلة ٠٠١ ، ٠٠٢ ، ٠٠٣ ، ٠٠٤ ، ٠٠٥ ، ٠٠٦ ، ٠٠٧ ، ٠٠٨ ، ٠٠٩ ، ٠١٠ ، ٠١١ ، ٠١٢ ، ٠١٣ ، ٠١٤ ، ٠١٥ ، ٠١٦ ، ٠١٧ ، ٠١٨ ، ٠١٩ ، ٠٢٠ ، ٠٢١ ، ٠٢٢ ، ٠٢٣ ، ٠٢٤ ، ٠٢٥ ، ٠٢٦ ، ٠٢٧ ، ٠٢٨ ، ٠٢٩ ، ٠٣٠ ، ٠٣١ ، ٠٣٢ ، ٠٣٣ ، ٠٣٤ ، ٠٣٥ ، ٠٣٦ ، ٠٣٧ ، ٠٣٨ ، ٠٣٩ ، ٠٤٠ ، ٠٤١ ، ٠٤٢ ، ٠٤٣ ، ٠٤٤ ، ٠٤٥ ، ٠٤٦ ، ٠٤٧ ، ٠٤٨ ، ٠٤٩ ، ٠٥٠ ، ٠٥١ ، ٠٥٢ ، ٠٥٣ ، ٠٥٤ ، ٠٥٥ ، ٠٥٦ ، ٠٥٧ ، ٠٥٨ ، ٠٥٩ ، ٠٦٠ ، ٠٦١ ، ٠٦٢ ، ٠٦٣ ، ٠٦٤ ، ٠٦٥ ، ٠٦٦ ، ٠٦٧ ، ٠٦٨ ، ٠٦٩ ، ٠٧٠ ، ٠٧١ ، ٠٧٢ ، ٠٧٣ ، ٠٧٤ ، ٠٧٥ ، ٠٧٦ ، ٠٧٧ ، ٠٧٨ ، ٠٧٩ ، ٠٨٠ ، ٠٨١ ، ٠٨٢ ، ٠٨٣ ، ٠٨٤ ، ٠٨٥ ، ٠٨٦ ، ٠٨٧ ، ٠٨٨ ، ٠٨٩ ، ٠٩٠ ، ٠٩١ ، ٠٩٢ ، ٠٩٣ ، ٠٩٤ ، ٠٩٥ ، ٠٩٦ ، ٠٩٧ ، ٠٩٨ ، ٠٩٩ ، ١٠٠ . يتألف كل منها من ٣ خانات ومن ثم نحتاج لاختيار العينة إلى ثلاثة أعمدة . نغمض أعيننا ونضع أصبعنا عشوائيا على أى نقطة في الجدول ولتكن هذه النقطة هي نقطة التقاء العمود ٨ والصف ٢١ حيث نجد العدد ٠٠٨ فيكون هذا الرقم هو الرقم المسلسل الأول في العينة . نقرأ ابتداء من هذا العدد رأسيا إلى أسفل ( مثلا ) بشكل هندسي ثابت مع إهمال الأعداد التي تزيد عن ٥٠٠ والأعداد التي يتكرر ظهورها . وإذا صادف أن انتهى العمود ٨ ولم يكتمل بعد الحجم المطلوب للعينة نتحول إلى ثلاثة أعمدة أخرى من أى من الجانبين الأيمن أو الأيسر ونقرأ من أعلى الجدول أو من أى مكان آخر في نفس الاتجاه السابق حتى نصل إلى الحجم المطلوب ونكون بذلك قد توصلنا إلى عينة عشوائية بسيطة قد تأخذ مفرداتها الأرقام المسلسلة الآتية :

٠٠٨ ١٤٢ ٠١٨ ١٤٣ ٤١٨ ٤٩١ ٢٤٠ ٠٢٠ ١٤٠ ٠٠٦  
١٠١ ٢٤٩ ٠١٦ ١٣٤ ٠٤٩ ٣١٩ ٤٨٣ ٢٣٢ ٠٥٢ ٠٢٩

**ملاحظة :** احتمال الحصول على عينة بهذه الطريقة يساوى احتمال الحصول على أى عينة أخرى من نفس الحجم ، وتؤخذ هذه الخاصة أحيانا كتعريف للعينة العشوائية البسيطة



### ( ١ - ٣ ) المتغيرات الإحصائية : STATISTICAL VARIABLES

في أي دراسة تطبيقية إحصائية ينبغي أن نحصل على بيانات عددية عن المتغيرات التي ندرسها . وتختلف طريقة تناولنا لهذه البيانات باختلاف تلك المتغيرات التي يمكن تقسيمها بصفة عامة إلى الأنواع الآتية :

أولا - المتغيرات الوصفية ( أو النوعية ) :

#### QUALITATIVE VARIABLES ( ATTRIBUTES )

وهي متغيرات لا يمكن قياس مفرداتها عدديا ولكن يمكن تقسيم هذه المفردات بحسب اشتراكها في صفة أو خاصية ما . ومن أمثلة هذه المتغيرات متغير اللون أو الجنس أو المهنة أو الخواص الوراثية . والجدول ( ١ - ١ ) هو مثال لبيانات وصفية فهو يعطي التوزيع التكرارى لألوان عيون عينة من ٨٠ فارا .

الجدول ( ١ - ١ )

التوزيع التكرارى لألوان عيون عينة من الفيران

اللون	التكرار
أسود	٤
ياقوتي	٢
رمادي	٧٤
المجموع	٨٠

الجدول ( ١ - ٢ )

ترتيب محكمين خمسة من الأشخاص

المرتبة	المرتبة الأولى	المرتبة الثانية
أ	٢	١
ب	١	٢
جـ	٤	٥
د	٣	٤
هـ	٥	٣

### ثانيا - المتغيرات الرتبية : RANKED VARIABLES

وهي أيضا متغيرات لا يمكن قياس مفرداتها عدديا ولكن يمكن تنظيم هذه المفردات بحسب ترتيب ما أو رتبة ما . ومن أمثلة هذه المتغيرات متغير النمو الذي

يمكن تقسيمه إلى ضعيف - عادي - مفرط ، كذلك المتغيرات التي تنتج من عملية التحكيم كما هو الحال مثلاً عندما يطلب إلى مجموعة من علماء النبات ترتيب عشرة نباتات من حيث شدة التلف الذى أصابها من مرض فطرى ، فيعطى كل منهم بحسب تقديره الترتيب (١) لأقل النباتات تأثراً بالمرض والترتيب (١٠) لأكثرها تأثراً ، أو حينما يطلب إلى مجموعة من الأطباء إبداء آرائهم في مشاهداتهم الميكروسكوبية عن مرض السرطان وترتيبها من حيث تزايد الورم السرطاني . الجدول ( ١ - ٢ ) يعطى الترتيب الذى رآه اثنان من المحكمين لخمسة من المتقدمين لشغل وظيفة ما .

#### ١١١٤ - المتغيرات العددية « الكمية » : QUANTITATIVE VARIABLES

وهى التى يمكن التعبير عنها بالأعداد العادية ( الحقيقية ) فيكون التغير فيها هو تغير من حيث المقدار أى يمكن تقسيم مفرداتها بحسب الأصغر والأكبر ، ونميز هنا بين النوعين الآتين :

##### ( أ ) المتغيرات المتصلة : CONTINUOUS VARIABLES

وهى تلك التى نحصل عليها عادة عن طريق القياس measurement مثل الطول والوزن والعمر ودرجة الحرارة . وفي هذه المتغيرات نستطيع أن نتصور أن المفردات يمكن أن تختلف عن بعضها بمقادير صغيرة صغراً لا نهائياً من الناحية النظرية . وإذا وقعت قيم متغير متصل بين عددين ١ ، ب فإنها تمثل بياناً بجميع نقاط قطعة مستقيمة محدودة بهذين العددين .

##### ( ب ) المتغيرات الوثابة : DISCRETE (OR MERISTIC) VARIABLES

وهى تلك التى نحصل عليها عادة عن طريق العد counting مثل عدد الذرية - عدد الخلفة - عدد ضربات القلب . ومجموعة قيم المتغير الوثاب قد تكون منتهية مثل { ٠ ، ١ ، ٢ ، ... ، ن } أو لا نهائية مثل { ٠ ، ١ ، ٢ ، ... } . وفي كلتا الحالتين تمثل بياناً بنقط منفصلة على خط مستقيم .

وفي المتغيرات العددية نميز من جهة أخرى بين المتغيرات العشوائية والمتغيرات غير العشوائية . والمتغيرات العشوائية هي متغيرات تخضع لمؤثرات عشوائية لا يمكن التحكم فيها تجريبيا ومن أمثلتها متغير درجات حرارة الجو ومتغير ضغط دم الإنسان أو تحصيله الدراسي ومتغير عدد ضربات قلب ضفدعة بعد معالجة ما ، وهذه جميعها تتأثر بعوامل عشوائية غير منظورة . أما المتغيرات غير العشوائية ( أو الرياضية ) فهي تلك التي لا تكون واقعة تحت تأثير عوامل عشوائية ولذلك يمكن قياسها بدقة أو بأخطاء صغيرة يمكن إهمالها .

ويهم الإحصاء بصفة خاصة بالمتغيرات العشوائية بل هي محور الدراسة فيه نظريا وتطبيقيا ولذلك سوف نتناولها بشيء من التفصيل هي وتوزيعاتها ونماذجها في الفصل الثالث من هذا الكتاب ليتيسر لنا التعامل معها بعد ذلك في الفصول التالية .

#### ( ١ - ٤ ) الأخطاء العشوائية ( أخطاء الصدفة ) RANDOM ERRORS

هناك أخطاء تخص المجال الذى يعمل فيه الباحث ولا تدخل في صميم الموضوعات الإحصائية كالأخطاء الناتجة عن عدم ضبط الأجهزة أو الأدوات المستخدمة في القياس أو عدم توفر الظروف الملائمة لإجراء التجربة كدرجة الحرارة أو الرطوبة ، أو الأخطاء الناتجة عن عدم ملائمة طريقة القياس ، أى عندما يكون هناك اختلاف بين التعريف النظرى للمتغير والتعريف الإجرائى المستخدم في عملية قياس هذا المتغير .

أما ما نهم به في الإحصاء فهي الأخطاء العشوائية ، وتظهر في الاختلافات التي نجدها في القيم العددية التي نحصل عليها عند القيام بقياسات متكررة لنفس الوحدة أو الشيء ، كما هو الحال مثلا عند قياس طول حشرة عدة مرات . ويتسبب في هذه الاختلافات عدد كبير من العوامل الثانوية التي نعجز عن حصرها أو تحديد مصدرها أو التنبؤ بها ونعجز عن حساب التأثير الضئيل الذى يحدثه كل منها على حدة . وبالتالي نعجز عن التحكم فيها تجريبيا ، وفي هذه الحال نحاول البحث عن وسائل تسمح بتقدير تأثيرها الكلى للإفادة من هذا التقدير في عملية التحليل الإحصائي

وحين تكون القياسات معرضة للأخطاء العشوائية فقط فإن أى قيمة  $s_r$  نحصل عليها من قياس عنصر منه تمثل بالمؤذج الآتى :

$$s_r = \bar{u} + \bar{x}_r$$

حيث  $\bar{u}$  هى القيمة الحقيقية للعنصر المقاس ،  $\bar{x}_r$  هى الخطأ فى  $s_r$  أى انحراف  $s_r$  عن القيمة الحقيقية  $\bar{u}$  .

إلا أنه على المدى البعيد تميل هذه الأخطاء إلى تعويض بعضها البعض بمعنى حدوث توازن بين الأخطاء الموجبة والأخطاء السالبة بحيث يقترب متوسط القياسات من القيمة الحقيقية للشيء الذى نقيسه ، ولذلك فإننا نفترض فى كثير من الحالات أن متوسط الأخطاء هو صفر على المدى البعيد .

#### **( ١ - ٥ ) قواعد تقريب الأعداد : RULES FOR ROUNDING**

كثيرا ما نلجأ إلى تدوين قياساتنا للمتغيرات العددية مقربة إلى عدد معين من الخانات . ولقد وجد أنه من المناسب الاتفاق على القواعد الآتية :

ليكن  $q$  هو الرقم الذى فى الخانة المراد التقريب إليها ،  $\bar{u}$  هو الرقم التالى من اليمين مباشرة للرقم  $q$  .

(أولا) إذا كان  $\bar{u} > ٥$  يبقى الرقم  $q$  كما هو .

(ثانيا) إذا كان  $\bar{u} < ٥$  يزداد الرقم  $q$  واحدا .

(ثالثا) إذا كان  $\bar{u} = ٥$  ،  $\bar{u}$  متبوعا بأرقام غير الصفر ، يزداد الرقم  $q$  واحدا .

أما إذا كان  $\bar{u} = ٥$  ،  $\bar{u}$  يقف وحده أو متبوعا بأصفار ، يبقى الرقم  $q$  كما هو إذا كان زوجيا ويزاد واحدا إذا كان فرديا . وهذه القاعدة الأخيرة من شأنها إحداث توازن بين الأعداد التى زيدت فى التقريب والأعداد التى نقصت ، خاصة إذا كان لدينا متتابعة طويلة من الأعداد المقربة .

مثال ( ١ - ٢ )

العدد	العدد مقرباً ( إلى عشرين عشريين )
أ ق	
٤٨,٣٢٤٦	٤٨,٣٢
٤٨,٣٢٦٠	٤٨,٣٣
٤٨,٣٢٥١	٤٨,٣٣
٤٨,٣٢٥٠	٤٨,٣٢
٤٨,٣١٥٠	٤٨,٣٢

مثال ( ١ - ٣ ) :

العدد	عدد الأرقام	العدد المقرب
	المعدية المطلوبة	
٢٦,٥٨	٢	٢٧
١٣٣,٧١٤٧	٥	١٣٣,٧١
١٣٣,٧١٤٧	٦	١٣٣,٧١٥
,٠٣٧٢٥	٣	,٠٣٧٢
,٠٣٧١٥	٣	,٠٣٧٢
١٨٣١٦	٢	١٨٠٠٠
١٨٣١٦	٣	١٨٣٠٠
١٧,٣٤٧٦	٣	١٧,٣

هذا مع ملاحظة أن آخر رقم في عدد مقرب ينبغي دائماً أن يكون معنوياً أى ينبغي أن يتضمن فترة تقع فيها القيمة الحقيقية للعدد ، وهذه الفترة تبدأ بنصف وحدة خطوة أسفل العدد المقرب المدون وتنتهى بنصف وحدة خطوة أعلاه فمثلاً العدد المقرب ٧,٨ يعنى أن قيمته الحقيقية تقع في الفترة بين ٧,٧٥ ، ٧,٨٥ أى  $٧,٨ \pm ٠,٠٥$  والعدد المقرب ٧,٨٠ يعنى أن قيمته الحقيقية تقع في الفترة بين

٧,٧٩٥ ، ٧,٨٠٥ أى  $7,80 \pm 0,05$  . والعدد المقرب ٣,٢٤ يعنى أن قيمته الحقيقية تقع فى الفترة بين ٣,٢٣٥ ، ٣,٢٤٥ أى  $3,24 \pm 0,005$  .

وحين نرغب فى تدوين قياس وحدة ما مقربة إلى عدد معين من الخانات نوجد هذا القياس بحيث يكون عدد الخانات الناتجة أكثر بواحد على الأقل من العدد المطلوب ثم نجرى التقريب . فمثلا لإيجاد خارج القسمة  $10 \div 7$  مقربا إلى ٣ خانات عشرية نقوم بعملية القسمة حتى نحصل على ٤ خانات ثم نقرب إلى ٣ خانات كالآتى :  $10 \div 7 = 1,428 = 1,43$  تقريبا .

## (١ - ٦) الدقة والضبط : ACCURACY AND PRECISION

الدقة هى تعبير عن مدى قرب قيمة نتجت عن قياس وحدة ما من القيمة الحقيقية لهذه الوحدة ، أما الضبط فهو تعبير عن مدى قرب القياسات المتكررة لوحدة ما من بعضها البعض تحت نفس الظروف . والإحصاء يهتم أساسا بالضبط لأن الضبط يتضمن الدقة طالما كانت أداة القياس غير متحيزة .

والدقة فى عدد مقرب يحكم عليها بدلالة النسبة المئوية للخطأ الذى يحتويه ، فمثلا نفرض أن عددا سجل على أنه ٨٩ تقريبا . إن هذا يعنى أن القيمة الحقيقية لهذا العدد تقع بين العددين ٨٨,٥ ، ٨٩,٥ وتكون القيمة المطلقة للحد الأقصى للخطأ هى ٠,٥ . وبالتالي تكون نسبة الخطأ هى :

$$\frac{0,5}{89} \times 100 = 0,56 \%$$

نفرض أن عددا آخر سجل على أنه ١٥,٥ تقريبا فتكون نسبة الخطأ هى :

$$\frac{0,5}{15,5} \times 100 = 3,22 \%$$

ونظرا لأن ٠,٥٦ أكبر من ٣,٢٢ نقول إن العدد ٨٩ أقل دقة من العدد ١٥,٥ أى أن العدد يكون أكثر دقة إذا استطعنا كتابته بعدد أكبر من الأرقام المعنوية .

أما الضبط فنحكم عليه بعدة طرق منها حجم الوحدة المستخدمة في القياس ، ففى قياس طول وحدة ما يكون القياس أكثر ضبطاً حين نستخدم مسطرة مدرجة بالملليمترات عنه حين نستخدم مسطرة مدرجة بالبوصات . وفى الإحصاء يتم التعبير عن الضبط فى كثير من الحالات بواسطة الانحراف المعيارى للقياسات المتكررة .

## PROBABILITY

( ١ - ٧ ) الاحتمال :

إن القرارات والأحكام والنتيوات التى نتخذها فى الإحصاء هى دائماً قرارات احتمالية بمعنى أننا لا نستطيع أن نجزم جزءاً باتناً بصحتها أو بخطئها ، فتأتى القرارات مصحوبة باحتمالات معينة للصواب أو الخطأ . ولعل هذا هو الذى أعطى الإحصاء قوته الكبرى وميزته عن الرياضيات ، إذ يتعرض للقضايا والفروض التى تعلن الرياضيات عجزها عن تناولها ، فبست فيها بأسلوب احتمالى هو على أية حال أفضل من ترك القضايا دون حل عملاً بالحكمة القائلة : « ما لا يدرك كله لا يترك كله » . ويهمنى فى مستهل دراستنا للإحصاء أن نتيين معنى الاحتمال من الناحية التطبيقية على الأقل .

إذا كان لدينا فرض ما فإننا نعين لصواب هذا الفرض العدد « ١ » ونعين لخطئه العدد « صفر » ومن ثم فإن أى عدد يقع بين الصفر والواحد يمكن أن يعبر عن درجة صواب أو خطأ هذا الفرض . إن الأعداد التى تبدأ بالصفر وتنتهى بالواحد هى التى نعبّر بها عن الاحتمالات ، فنقول مثلاً إن احتمال ظهور البترول فى منطقة ما هو ٠,٦ أو إن احتمال الحصول على نباتات حمراء من مجموعة معينة من بذور حنك السبع هو ٠,٤ وهكذا ... ويؤخذ الاحتمال بداهة كمقياس لدرجة اعتقادنا أو شدة اقتناعنا فى صحة فرض ما أو فى وقوع حدث ما . وليس من المناسب أن نعتبر هذا تعريفاً للاحتمال لأنه يخضع لذاتية المشاهد وخبرته ، ومع ذلك فكثيراً ما نجد أنفسنا مضطرين إلى أن نقدر احتمال حدث ما بالبداهة أو الخبرة أو أى عوامل ذاتية أخرى وذلك حين لا تتوفر عوامل مباشرة تكفى لحساب احتمال الحدث .

والتعريف الدقيق للاحتيال يتألف من مجموعة من المسلمات المصاغة صياغة رياضية ، على أن الدراسات التطبيقية لا تحتاج إلى هذا التعريف وإنما تكتفى بتعريفين في مستوى أقل ، فيعتبر الأول أن الاحتمال هو « نسبة » ويعتبر الثاني أن الاحتمال هو « تكرار نسبي » كما هو موضح بعد .

#### ( أ ) التعريف التقليدي : CLASSICAL DEFINITION

إذا أسفرت تجربة عن  $n$  من النواتج أو الحالات المتساوية الإمكانيات وكان الحدث  $A$  يقع في  $m$  من هذه النواتج فإن احتمال وقوع هذا الحدث يساوى  $\frac{m}{n}$  أى يساوى النسبة بين عدد الحالات التى يمكن أن يقع فيها الحدث والعدد الكلى للحالات التى يمكن أن تسفر عنها التجربة .

فمثلا إذا ألقينا حجرة نرد منتظمة عشوائيا تكون النواتج الممكنة للتجربة هى الأعداد الست ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ وهذه النواتج متساوية الإمكانيات لأنه لا يوجد لدينا ما يجعلنا نتوقع ظهور أحدها دون الآخرين وعلى ذلك فإن احتمال الحدث « عدد فردى » مثلا هو  $\frac{3}{6} = \frac{1}{2}$  . كذلك احتمال سحب « صورة » من مجموعة محكمة الخلط من ورق اللعب  $\frac{12}{52} = \frac{3}{13}$  .

#### ( ب ) التعريف الإحصائى : STATISTICAL DEFINITION

ينطلق هذا التعريف من فكرة التكرار النسبى ومن ظاهرة اكتشفت بالملاحظة والتجريب تعرف بظاهرة الانتظام الإحصائى Statistical regularity ومجملها أنه إذا كررت تجربة مرات كثيرة تحت نفس الظروف ( مثل زراعة بذرة من نبات حنك السبع ) فإن التكرار النسبى لحدث ما متعلق بهذه التجربة ( مثل ظهور اللون الأحمر ) يقترب من عدد ثابت كلما زاد عدد مرات إجراء التجربة ، ويؤخذ هذا العدد كتقدير لاحتمال ذلك الحدث . ولذلك يعرف احتمال حدث ما بأنه نهاية متتابعة من التكرارات النسبية لوقوع هذا الحدث . فمثلا إذا ألقينا قطعة نقود منتظمة عشوائيا عشر مرات ثم مائة ثم ألف ثم ... فإن التكرار النسبى لظهور



الصورة يقترب من العدد  $\frac{1}{p}$  كلما زاد عدد مرات إلقاء القطعة . وهنا نقول إن احتمال الحدث « ظهور الصورة » هو  $\frac{1}{p}$  . ويفهم من هذا أنه إذا رميت قطعة نقود عددا كبيرا من المرات فإن نسبة عدد المرات التي تظهر فيها الصورة إلى العدد الكلى لمرات رمي القطعة هو  $\frac{1}{p}$  ( على المدى الطويل ) .

ويوحى التعريف الإحصائي بطريقة مناسبة لإيجاد احتمالات الأحداث تجريبيا ، فإذا أردنا مثلا إيجاد احتمال وجود وحدة معينة من الوحدات التي ينتجها مصنع ما ، نسحب عينة عشوائية كبيرة من هذه الوحدات ونحسب التكرار النسبي للوحدات المعنية فنحصل على تقدير للاحتمال المطلوب ، ويمكن بعد ذلك اختبار دقة هذا التقدير بالطرق الإحصائية التي سندرسها بعد .

أما التعريف التقليدي فيصلح لإيجاد الاحتمالات نظريا في الحالات التي يتوفر فيها شرط تساوى الإمكانيات كما في المثال الآتى .

مثال ( ١ - ٤ ) :

في مجموعة من ١٠٠ رجل علم أن ١٢ منهم يلبسون نظارات ، ٨ منهم لا يلبسون نظارات ولكن يحتاجون إليها . إذا اخترنا رجل واحد من هذه المجموعة عشوائيا فإن :

- ( أ ) احتمال أن يكون الرجل لابساً نظارة = ٠,١٢
- ( ب ) احتمال أن يكون الرجل « لا يلبس نظارة ولكن يحتاج إليها » = ٠,٠٨
- ( ج ) احتمال أن يكون الرجل « لا يلبس نظارة ولا يحتاج إليها » = ٠,٨٠

امتداد للتعريف التقليدي :

إن التعريف التقليدي للاحتمال الذى سبق تقديمه يتناول التجارب التي تسفر عن عدد متبى من النواتج ، إلا أنه مع بعض التعديل يمكن أن يمتد ليشمل التجارب التي تسفر عن عدد غير متبى من النواتج كذلك التي يكون فيها لقضاء التجريب مقياس هندسى كالطول أو المساحة أو الحجم . فإذا كان هذا القضاء يتألف من

منطقة محددة ١ وكانت المنطقة ب جزءا من المنطقة ١ واخترنا عشوائيا نقطة من المنطقة ١ فإن احتمال الحدث ١ وقوع النقطة في المنطقة ب ١ يعرف بالنسبة :

$$\frac{\text{مقياس ب}}{\text{مقياس ١}}$$

وذلك مع الاحتفاظ بفرض تساوى الإمكانيات . ويلاحظ هنا أن تعبير العشوائية يعنى أن احتمال وقوع نقطة في جزء من فضاء التجريب يتناسب مع مقياس هذا الجزء وأن هذا الاحتمال مستقل عن شكل المنطقة وموضعها .

مثال ( ١ - ٥ )

اختيرت نقطة عشوائيا من داخل دائرة ١ نصف قطرها ٣ سم . ما احتمال ألا يزيد بعد هذه النقطة عن مركز الدائرة عن ٢ سم ؟  
الحل :

يقع الحدث المطلوب إذا وقعت النقطة المختارة داخل دائرة ب لها نفس مركز الدائرة ١ ونصف قطرها ٢ سم .

$$\therefore \text{الاحتمال المطلوب} = \frac{\text{مقياس الدائرة ب}}{\text{مقياس الدائرة ١}} = \frac{٤ \times \pi}{٩ \times \pi} = \frac{٤}{٩}$$

مثال ( ١ - ٦ )

اختيرت نقطة عشوائيا على القطعة المستقيمة ١ = { ٣ : ١٠ ≤ ٣ ≤ ٠ }  
ما احتمال وقوع النقطة في القطعة المستقيمة ب = { ٣ : ٨ ≤ ٣ ≤ ٢ }  
الحل :

$$\text{الاحتمال المطلوب} = \frac{\text{طول القطعة ب}}{\text{طول القطعة ١}} = \frac{٦}{١٠} = ٠,٦$$

## اصطلاحات وتعاریف :

(١) الرمز ل ( ا ) یعنی احتمال وقوع الحدث ا ، ولاحظ أن

$$١ \leqslant ل ( ا ) \leqslant ٠$$

وحین ل ( ا ) = ٠ نقول إن الحدث ا هو حدث مستحيل ،

وحین ل ( ا ) = ١ نقول إن الحدث ا هو حدث مؤكد .

(٢) الرمز ل ( ا ، أو ا٢ )

يعنی احتمال وقوع واحد على الأقل من الحدثين ا١ ، ا٢ أى وقوع ا١ فقط أو ا٢ فقط أو ا١ ، ا٢ معا .

(٣) الرمز ل ( ا١ و ا٢ )

يعنی احتمال وقوع الحدثين ا١ ، ا٢ معا أو بالتتابع .

(٤) الرمز ل ( ا١ | ا٢ )

يعنی احتمال وقوع الحدث ا١ بشرط أن يكون الحدث ا١ قد وقع فعلا ويسمى هذا الاحتمال بالاحتمال الشرطى للحدث ا١ بالنسبة للحدث ا١ .

مثال ذلك احتمال اختيار طالب من مدرسة ما بشرط أن يكون من الرياضيين .

(٥) يقال للحدثين ا١ ، ا٢ إنهما متافيان إذا كان وقوع أحدهما يمنع وقوع الآخر وبذلك لا يمكن أن يقع الحدثان معا ، أى يكون ل ( ا١ و ا٢ ) = ٠ فمثلا فى حالة ولادة طفل يكون الحدث « المولود ذكر » والحدث « المولود أنثى » حدثين متافيين ( لا يمكن أن يقعا معا ) .

(٦) يقال للحدثين ا١ ، ا٢ إنهما مستقلان إذا كان احتمال وقوع أيهما لا يتأثر بوقوع أو عدم وقوع الآخر ، أى إذا كان :

$$L(A, B) = L(A, B), \quad L(A, B|C) = L(A, B|C)$$

فمثلا إذا ألقينا قطعتين من العملة عشوائيا فإن ما يظهر على أى منهما ( صورة أو كتابة ) يكون مستقلا عما يظهر على الأخرى ( إلا إذا كانت القطعتان مربوطتين معا بخيط مثلا ) . كذلك اختيار طالب من كلية العلوم واختيار طالب من كلية الإنسانيات هما حدثان مستقلان .

### توافقات الاحتمال :

القاعدتان الآتيتان تسهلان حساب الاحتمالات في كثير من الحالات ويمكن إثباتهما رياضياً :

### أولاً - قاعدة الجمع :

$$L(A, B \text{ أو } C) = L(A, B) + L(A, C) - L(A, B \text{ و } C)$$

أى أن احتمال وقوع أحد الحدثين  $A, B$  أو كلاهما يساوى احتمال وقوع الأول + احتمال وقوع الثانى - احتمال وقوعهما معا .

### حالة خاصة :

إذا كان الحدثان  $A, B$  متنافيين فإنه حسب التعريف (٥) تصبح قاعدة الجمع كالآتى :

$$L(A, B \text{ أو } C) = L(A, B) + L(A, C)$$

وتسمى هذه القاعدة بقاعدة الجمع للأحداث المتنافية ، ويمكن تعميمها كالآتى :

إذا كانت  $A, B, C, \dots$  ، أن أحداثا متنافية فإن :

$$L(A, B \text{ أو } C \text{ أو } \dots \text{ أو } N) = L(A) + L(B) + L(C) + \dots + L(N)$$

### مثال ( ١ - ٧ )

في مجموعة من ١٠٠ طالب رسب ١٢ في الرياضيات ورسب ١٥ في الفيزياء ورسب ٨ في كلتا المادتين . إذا اختير طالب واحد عشوائيا من هذه المجموعة فما احتمال أن يكون راسبا في الرياضيات أو في الفيزياء ؟  
الحل :

نلاحظ أن الرسوب في أى مادة لا يمنع الرسوب في المادة الأخرى ، أى أن الحدثين غير متنافيين وحسب قاعدة الجمع نجد أن :  
ل (راسب في الرياضيات أو راسب في الفيزياء) =  $0,12 + 0,15 - 0,08 = 0,19$ .

### مثال ( ١ - ٨ ) :

ألقي حجر نرد منتظم عشوائيا مرة واحدة . نجد أن :  
(١) احتمال ظهور ٥ أو ٦ = ل (٥ أو ٦)  
ل = ل (٥) + ل (٦) (حدثان متنافيان)  
 $\frac{1}{6} = \frac{1}{6} + \frac{1}{6}$   
(ب) احتمال ظهور عدد فردى = ل (١ أو ٣ أو ٥)  
(٣ أحداث متنافية)  $\frac{1}{2} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} =$

نتيجة :

إذا كانت ١، ٢، ... ، أن هي جميع الأحداث الممكنة في تجربة ما وكانت هذه الأحداث متنافية فإن مجموع احتمالات هذه الأحداث يساوى الواحد الصحيح ، أى أن :

$$ل (١) + ل (٢) + ... + ل (أن) = ١$$

لاحظ تحقق هذه النتيجة في المثال ( ١ - ٤ ) السابق .

## ثانيا - قاعدة الضرب :

$$\begin{aligned} & \text{ل (أ, ب) = ل (أ) . ل (ب | أ) ، ل (أ | ب) . ل (ب) \neq \\ & \text{أو ل (أ) = ل (أ) . ل (ب | أ) ، ل (أ | ب) . ل (ب) \neq \end{aligned}$$

أى أن احتمال وقوع حدثين معا يساوى احتمال أحدهما مضروباً فى الاحتمال الشرطى للآخر بالنسبة للأول .

## حالة خاصة :

إذا كان الحدثان أ ، ب مستقلين فإنه حسب (٦) تصبح القاعدة كالآتى :

$$\text{ل (أ, ب) = ل (أ) . ل (ب | أ) . ل (أ) . ل (ب)}$$

وتسمى هذه النتيجة بقاعدة الضرب للأحداث المستقلة ، ويمكن تعميمها كالآتى :

إذا كانت أ ، ب ، ... ، أن أحداثاً مستقلة فإن :

$$\text{ل (أ, ب, و ... وان) = ل (أ) . ل (ب) . ل (و) ... ل (ان)}$$

## مثال ( ٩ - ٩ ) :

رمى حجراً نرد منتظمان وتميزان عشوائياً مرة واحدة . نجد أن :

$$\begin{aligned} & \text{( أ ) احتمال ظهور ٥ على القطعة الأولى و ٦ على القطعة الثانية} \\ & \text{ل ( ٥ ٦ ) = } \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \quad \text{( حدثان مستقلان )} \\ & \text{( ب ) احتمال ظهور ٦ على القطعة الأولى و ٥ على القطعة الثانية} \\ & \text{ل ( ٦ ٥ ) = } \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \quad \text{( حدثان مستقلان )} \\ & \text{( ج ) احتمال ظهور ٥ على إحدى القطعتين و ٦ على القطعة الأخرى} \\ & \text{ل ( ٥ ٦ ) + ل ( ٦ ٥ ) = } \frac{1}{36} + \frac{1}{36} = \frac{2}{36} = \frac{1}{18} \end{aligned}$$

مثال ( ١ - ١٠ ) :

ثلاث مجموعات من الأطفال تتألف الأولى من ( ٣ بنات ، ولد واحد )  
وتتألف الثانية من ( بنتين ، ولدين ) وتتألف الثالثة من ( بنت واحدة ، ٣  
أولاد ) . اختير طفل واحد عشوائياً من كل مجموعة . ما احتمال أن يكون الثلاثة  
الأطفال المختارون عبارة عن بنت واحدة وولدين ؟

الحل :

يقع الحدث المطلوب بإحدى الطرق الثلاث الآتية :

( بنت - ولد - ولد ) أو ( ولد - بنت - ولد ) - أو ( ولد - ولد - بنت )  
احتمال الطريقة الأولى  $= \frac{3}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{6}{64}$  ( أحداث مستقلة )

احتمال الطريقة الثانية  $= \frac{2}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{4}{64}$  ( أحداث مستقلة )

احتمال الطريقة الثالثة  $= \frac{1}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{2}{64}$  ( أحداث مستقلة )

وبما أن هذه الطرق الثلاثة متنافية فإن احتمال أن يكون الأطفال المختارون عبارة  
عن بنت واحدة وولدين هو مجموع احتمالات هذه الطرق ، أى :

$$\frac{12}{64} = \frac{6}{64} + \frac{4}{64} + \frac{2}{64}$$

مثال ( ١ - ١١ )

صندوق به ٧ كرات حمراء و ٣ كرات بيضاء . سحب كرتان عشوائياً  
الواحدة بعد الأخرى دون إرجاع . احسب احتمالات الأحداث الآتية :  
( أ ) أن تكون كلا الكرتين حمراء . ( ب ) أن تكون كلا الكرتين بيضاء .  
( ج ) أن تكون إحدى الكرتين حمراء والأخرى بيضاء .

الحل :

( أ ) احتمال أن تكون الكرة الأولى حمراء =  $\frac{2}{11}$

احتمال أن تكون الكرة الثانية حمراء =  $\frac{1}{9} = \frac{2}{9}$

( هذا احتمال شرطى مع ملاحظة أنه بعد السحبة الأولى يبقى في الصندوق ٩ كرات منها ٦ حمراء ) .

احتمال أن تكون كلا الكرتين حمراء =  $\frac{2}{11} \times \frac{2}{9} = \frac{4}{99}$

وذلك باستخدام قاعدة الضرب وملاحظة أن الاحتمال  $\frac{2}{9}$  هو احتمال شرطى فهو احتمال ظهور كرة حمراء في السحبة الثانية بشرط أن تكون السحبة الأولى حمراء .  
( ب ) بنفس المنطق نجد أن :

احتمال أن تكون كلا الكرتين بيضاء =  $\frac{2}{11} \times \frac{1}{9} = \frac{2}{99}$

(جـ) احتمال أن تكون الكرة الأولى حمراء والثانية بيضاء =  $\frac{2}{11} \times \frac{1}{9} = \frac{2}{99}$

واحتمال أن تكون الكرة الأولى بيضاء والثانية حمراء =  $\frac{1}{11} \times \frac{2}{9} = \frac{2}{99}$

احتمال أن تكون إحدى الكرتين حمراء والأخرى بيضاء =  $\frac{2}{99} + \frac{2}{99} = \frac{4}{99}$

( حدثان متافيان )

## ( ٩ - ٨ ) النماذج الرياضية : MATHEMATICAL MODELS

النموذج لظاهرة ما هو وسيلة لبيان الميكل العالم لتركيب هذه الظاهرة أو للتعبير عن نظرية أو وضع يؤخذ كمولد لما نلمسه من مشاهدات عن هذه الظاهرة . وفي التحليل الإحصائي عادة ما توضع النماذج في صورة رياضية وإن كان هناك صور أخرى بيانية كالخرائط الهندسية والجغرافية والجيولوجية .



والنموذج الرياضي هو تجريد لنموذج فيزيائي حيث نحل محل الأشياء والقوى والأحداث رموز تعبر عن متغيرات وبارامترات وثوابت . ويمكن تقسيم النماذج الرياضية إلى ثلاثة أنواع هي : النماذج التحديدية - النماذج الإحصائية - نماذج العمليات العشوائية .

#### (أ) النماذج التحديدية : DETERMINISTIC MODELS

هي تلك النماذج الرياضية المعتادة التي تستخدم في مختلف المجالات العلمية والتي تعبر عن العلاقة الدالية بين المتغيرات على هيئة قوانين أو معادلات جبرية أو تفاضلية أو تكاملية أو مصفوفية يمكن اشتقاقها نظرياً دون اللجوء إلى التجريب ، كقوانين نيوتن للحركة .

غير أنه عند التحقق من صحة هذه القوانين تجريبياً نتعرض إلى عدة مصادر للخطأ منها الخطأ في القياس measurement error ومنها خطأ المعادلة equation error أى خطأ استخدام معادلة غير مناسبة ، ومنها الأخطاء العشوائية السابق الإشارة إليها في البند ( ١ - ٤ ) . إن هذه الأخطاء لا تدخل في اعتبار النموذج التحديدي ويتطلب تقييمها تحويل هذا النموذج إلى نموذج إحصائي .

#### (ب) النماذج الإحصائية : STATISTICAL MODELS

النموذج الإحصائي هو تعبير رياضي يشتمل على واحد أو أكثر من المركبات العشوائية بالإضافة إلى المتغيرات والبارامترات والثوابت التي يشتمل عليها النموذج التحديدي . ويمكن اشتقاق نموذج إحصائي من نموذج تحديدي بإضافة صريحة للمركبات العشوائية . فمثلاً القانون الشهير الذي يربط المسافة والزمن لجسيم يسير بسرعة ابتدائية  $c$  وعجلة منتظمة  $a$  يأخذ الصورة :

$$s = ct + \frac{1}{2}at^2$$

وهذا نموذج تحديدي . إذا أجرينا تجربة عدة مرات وكانت  $s$  ترمز إلى

المسافة المقطوعة في التجربة الرائية ، نستطيع أن ندخل العامل العشوائي صراحة كالآتي ( على أساس أن ع ، ن ، ح ثوابت ) :

$$ف_r = ع_r + ن_r + \frac{1}{4} ح_r + \frac{1}{4} خ_r$$

$$لوف_r = ف_r + خ_r$$

حيث ف هي القيمة الحقيقية للمسافة كما تحسب من النموذج التحديدي ، خ<sub>r</sub> هي انحراف المسافة الناتجة في التجربة الرائية عن المسافة الحقيقية ف .

وسنلقى أمثلة كثيرة للنماذج الإحصائية خاصة عند دراسة تحليل التباين وتحليل الانحدار .

على أنه ليس من الضروري أن يعبر النموذج الإحصائي عن نموذج فيزيائي بالوضوح الظاهر في القانون السابق بل يمكن اختيار نموذج إحصائي مناسب اعتماداً على ما نتصوره من مصادر تؤثر فيما نشاهده من الاختلافات في البيانات الناتجة عن الظواهر التي ندرسها .

### (ج) نماذج العمليات العشوائية : STOCHASTIC PROCESS MODELS

هذا النوع من النماذج قد يشتمل على عوامل عشوائية مماثلة لتلك التي في النماذج الإحصائية ولكنه بالإضافة إلى ذلك يشتمل على عملية عشوائية معينة تدخل في بناء النموذج وتصف الظاهرة على أساس احتمالي وليس على أساس تحديدي . ومن هذه النماذج النموذج II الذي ستقدمه في فصل تحليل التباين بالبند ( ٨ - ١٤ ) .

## تمارين (١)

١ - الآتي بيان الأعمار بالسنوات للأطفال الذين ذهبوا إلى إحدى المستشفيات للعلاج في أحد الأيام وعددهم ٢٤٠ .

(أ) أعط هذه الأعمار أرقاماً متسلسلة من ٠٠١ إلى ٢٤٠

(ب) استخدم جدول الأعداد العشوائية لاستخراج عينة عشوائية حجمها ٣٠ (اذكر رقم العمود ورقم الصف اللذين بدأت بهما) .

(ج) أوجد الوسط الحسابي لأعمار أفراد هذه العينة .

(د) كرر الخطوتين (ب) ، (ج) عدة مرات وقارن بين الأوساط الحسابية للعينات الناتجة وامن الوسط الحسابي لأعمار الأطفال جميعاً .

٢	٨	١٠	٣	١٠	٧
٥	١٢	١٣	٧	١٤	٥
٥	٢	١١	٦	١١	٧
٥	١	٥	٢	٢	٦
٩	٢	٣	٤	٥	١٣
٤	٤	١	١	٣	٢
١٣	١	١٥	٦	١	٢
٨	٥	٢	٦	٥	٥
٥	١٠	١٤	٩	٢	٨
١٤	٧	٦	١٤	٣	٤
٥	٨	٢	١	٨	٣
١	١	٤	٢	١	٤
١٣	١٥	١	١	٢	٩
٨	٥	٣	٤	٥	٣
١٤	٣	٧	١١	٦	٢

1	3	8	10	9	11
3	7	1	1	1	8
10	8	1	0	0	2
7	1	2	9	12	1
2	12	0	7	10	2
2	2	11	1	2	10
11	10	0	7	9	7
7	2	2	2	10	1
2	0	2	1	2	13
12	10	13	3	1	10
3	2	9	2	3	7
1	2	0	7	8	2
7	12	13	3	11	1
0	9	3	11	1	12
10	1	0	2	2	10
2	12	9	3	7	3
9	10	1	9	13	12
2	0	12	0	1	0
13	12	3	7	7	12
3	2	3	1	3	3
7	10	8	12	9	2
2	0	0	9	2	9
7	1	2	3	2	2
12	0	12	1	1	1
3	12	7	7	10	8

٢ - قرب الأعداد الآتية إلى درجة الدقة المشار إليها .  
 ٦٥,٤ إلى أقرب وحدة ، ٠,٠٦٣٥ إلى أقرب جزء من ألف  
 ١٤٤,٥ إلى أقرب وحدة ، ٢,٥٠٠١ إلى أقرب وحدة  
 ٣,٥٤٢ إلى أقرب جزء من مائة ، ٢٤,٦٢ إلى رقمين معنويين .

٣ - اجمع الأعداد ٨,٣٤ ، ٦,١٢ ، ٧,٥٤ ، ٦,٥٦ ، ٦,٤٧ ، ٧,٢٨ ،  
 ٧,١٧

( أ ) مباشرة (ب) بعد تقريب كل منها إلى جزء من عشرة .

٤ - صندوق به ٦ كرات حمراء و ٤ كرات صفراء . سحبنا منه عشوائياً كرتين  
 الواحدة بعد الأخرى دون إعادة .

- ( أ ) احسب احتمال أن تكون كلا الكرتين حمراء .
- (ب) احسب احتمال أن تكون كلا الكرتين صفراء .
- (ج) احسب احتمال أن تكون واحدة حمراء وواحدة صفراء .

٥ - مجموعتان من الأطفال تتألف الأولى من ( ٣ بنات ، ولدين ) وتتألف الثانية  
 من ( بنتين ، ٣ أولاد ) . اختر طفل واحد عشوائياً من كل مجموعة . ما  
 احتمال أن يكون الطفلان المختاران عبارة عن بنت واحدة وولد واحد ؟ .



## الفصل الثاني

### التوزيعات التكرارية

#### FREQUENCY DISTRIBUTIONS

إن البيانات التي نحصل عليها من عينة ما عن متغير ما تكون عبارة عن عدد من القيم أو القراءات مسجلة كيفما اتفق ، ولذلك تسمى عادة بالبيانات الخام raw data . وأول ما نفعله بهذه البيانات هو تنظيمها وتلخيصها في صورة مركزة عادة ما تكون على هيئة توزيعات تكرارية موضوعة في جداول مناسبة ، وكثيراً ما نقوم بتمثيلها بيانياً . إن هذا التنظيم يجعل البيانات أكثر طواعية للدراسة والتحليل وقد يكشف عن بعض الصفات البارزة أو الخصائص العامة التي لا تظهر في القراءات قبل تنظيمها . ولا مفر لأى دارس للإحصاء من أن يكون على دراية بأمور أساسية ثلاثة هي :

( أ ) كيفية إنشاء الجداول التكرارية .

( ب ) كيفية تمثيل التوزيعات بيانياً .

( ج ) كيفية وصف التوزيعات وصفاً موضوعياً .

وهذا ما نتناوله هنا بالتلخيص المركز عن طريق الأمثلة ، على أساس أن القارئ سبق له دراستها .

#### FREQUENCY TABLES

#### ( ٢ - ١ ) الجداول التكرارية :

#### ( ٢ - ١ - ١ ) جدول التوزيع التكرارى البسيط :

مثال ( ٢ - ١ ) :

الأعداد الآتية تعبر عن النسب المئوية للكاربون الذى وجد في عينة عشوائية حجمها ٢٥ في نوع من الفحم .

٨٠ ٨١ ٨٦ ٨٧ ٨٥ ٨٦ ٨٧ ٨٢ ٧٧ ٨٠ ٨٣ ٨٤ ٨٦  
٨٧ ٧٧ ٨٦ ٧٩ ٧٩ ٨٢ ٧٨ ٨٥ ٨٦ ٨٤ ٨٥ ٨٣

كون الجدول التكرارى البسيط وجدول التكرارات المتجمعة المفوية .

الحل :

نشئ جدولاً كالجدول (٢ - ١) التالى حيث يشتمل العمود الأول منه على القيم المختلفة للمتغير مرتبة ترتيباً تصاعدياً ( أو تنازلياً ) ويشتمل العمود الثانى منه على عدد من الشرط أمام كل قيمة بالعمود الأول تحصى عدد مرات وجود هذه القيمة بالبيانات الخام . أما العمودان الثالث والرابع فأمرهما واضح وكذلك بالنسبة للجدول (٢ - ٢) .

الجدول (٢ - ١)

التكرارات المتجمعة والمتجمعة المفوية

الحدود العليا	التكرار المتجمع	التكرار المتجمع %
٧٧ >	٢	٨
٧٨ >	٣	١٢
٧٩ >	٥	٢٠
٨٠ >	٧	٢٨
٨١ >	٨	٣٢
٨٢ >	١٠	٤٠
٨٣ >	١٢	٤٨
٨٤ >	١٤	٥٦
٨٥ >	١٧	٦٨
٨٦ >	٢٢	٨٨
٨٧ >	٢٥	١٠٠

الجدول (٢ - ١)

التكرارات والتكرارات النسبية

النسب المفوية للكاريون في عينة من القمح

م،	الشرط	ك،	ح،
٧٧	//	٢	٠,٠٨
٧٨	/	١	٠,٠٤
٧٩	//	٢	٠,٠٨
٨٠	//	٢	٠,٠٨
٨١	/	١	٠,٠٤
٨٢	//	٢	٠,٠٨
٨٣	//	٢	٠,٠٨
٨٤	//	٢	٠,٠٨
٨٥	////	٣	٠,١٢
٨٦	<del>////</del>	٥	٠,٢٠
٨٧	////	٣	٠,١٢
المجموع		٢٥	١,٠٠



## ملاحظات :

- (١) س , ترمز إلى القيم المختلفة للمتغير .  
 (٢) ك , ترمز إلى تكرار القيمة س , أى إلى عدد مرات وجود هذه القيمة بالبيانات الخام .  
 (٣) ح , ترمز إلى التكرار النسبي للقيمة س , أى خارج قسمة العدد ك , على حجم التوزيع وهو هنا ٢٨ . وهذه التكرارات النسبية تؤخذ كتقديرات للاحتالات تحت شروط معينة منها عشوائية العينة وكبر حجمها .

(٤) تسمى مجموعة الأزواج المرتبة

$$\{ (٣ , ٨٧) , (١ , ٧٨) , (٢ , ٧٧) \}$$

بالتوزيع التكرارى للمتغير س في العينة ، وهذه المجموعة تشكل العمودين الأول والثالث من الجدول (٢ - ١) .

(٢ - ١ - ٢) جدول التوزيع التكرارى المجمع في فئات :

حين تشتمل البيانات على عدد كبير من قيم متغير عددي ، يفضل تجميع هذه القيم في فئات فتوضع كل مجموعة من القيم المقاربة في فئة خاصة ، ويراعى هنا ألا يكون عدد هذه الفئات كبيراً فتنتفي الحكمة أو الفائدة من عملية التجميع ، وألا يكون عددها صغيراً فتضيع معالم التوزيع ويفقد الكثير من تفاصيله .

مثال (٢ - ٢) :

قيست أطوال محيطات الرؤوس بالمليمترات لعينة حجمها ٤٠ من الحمام المنزلى

فوجدت كما يلي :

١١,٦ ١١,٨ ١١,٨ ١٢,٢ ١٢,٣ ١١,١ ١١,٩ ١١,٨ ١٢,٩ ١٢,٢  
 ١١,١ ١٠,٥ ١١,٢ ١٣,٣ ١١,٩ ١١,٦ ١١,٢ ١١,٣ ١١,٥ ١٠,٧  
 ١١,٦ ١٠,٩ ١٠,٢ ١١,٩ ١١,٠ ١٠,٨ ١٠,٧ ١٠,٤ ١١,٩ ١٢,١  
 ١١,١ ١١,٣ ١١,٨ ١١,٧ ١٢,٤ ١٢,٠ ١٠,٧ ١٠,٤ ١١,٦ ١٠,٨

جمع هذه البيانات في توزيع تكرارى ذى فئات وأوجد توزيع التكرارات المتجمعة النسبية المئوية .

**الحل :**

المدى = أكبر قيمة للمتغير - أصغر قيمة للمتغير

$$= 13,3 - 10,2 = 3,1$$

إذا رأينا أن نأخذ حوالى ١٠ فئات يكون طول كل فئة  $3,1 \div 10 = 0,30$  تقريباً .

وعلى أساس أن الأطوال قيمت لأقرب جزء من عشرة من المليمتر سنعتبر أنه إذا كان س هو العدد الذى سجلناه لطول محيط رأس حمامة فإن الطول الحقيقي لهذا الرأس يقع بين العددين  $0,05 \pm$  فمثلاً أصغر عدد مسجل هو ١٠,٢ وإذن الطول الحقيقي لمحيط رأس أصغر حمامة في العينة يقع بين العددين  $10,2 \pm 0,05$  أى بين العددين ١٠,١٥ ، ١٠,٢٥ .

نأخذ العدد ١٠,١٥ كحد أدنى للفئة الأولى . وحيث أننا اخترنا أن يكون طول الفئة ٠,٣٠ فإن الحد الأعلى لهذه الفئة يكون  $10,15 + 0,30 = 10,45$  ويكون هذا العدد نفسه هو الحد الأدنى للفئة الثانية التي ينبغي أن يكون حدها الأعلى  $10,45 + 0,30 = 10,75$  ونستمر في إنشاء الفئات التالية بنفس الطريقة حتى نصل إلى فئة تغطى أكبر عدد في البيانات المعطاة وهو ١٣,٣ — انظر العمود الأول من الجدول (٢ - ٣) الآتي . ويلاحظ أن هذا الأسلوب في تكوين الفئات للقياسات المقربة يسمح بأن يكون لكل عدد في البيانات المعطاة مكان وحيد في إحدى هذه الفئات . وبالنسبة للمأ العمود الثالث نمر على الأعداد المعطاة واحداً واحداً ونضع شرطة أمام الفئة التي يدخل فيها .

الجدول (٢ - ٣)  
التكرارات والتكرارات النسبية لمعطيات الرؤوس  
بالمليمترات لمعينة الخمام المنزلى

ح ر	ك ر	الشرط	م ر	القيمات
٠,٠٧٥	٣	///	١٠,٣	١٠,٤٥-١٠,١٥
٠,١١٠	٤	////	١٠,٦	١٠,٧٥-١٠,٤٥
٠,١٠٠	٤	////	١٠,٩	١١,٠٥-١٠,٧٥
٠,١٧٥	٧	//###	١١,٢	١١,٣٥-١١,٠٥
٠,١٢٥	٥	++++	١١,٥	١١,٦٥-١١,٣٥
٠,٢٢٥	٩	////###	١١,٨	١١,٩٥-١١,٦٥
٠,١٠٠	٤	////	١٢,١	١٢,٢٥-١١,٩٥
٠,٠٥٠	٢	//	١٢,٤	١٢,٥٥-١٢,٢٥
٠,٠٠٠	٠	-	١٢,٧	١٢,٨٥-١٢,٥٥
٠,٠٢٥	١	- /	١٣,٠	١٣,١٥-١٢,٨٥
٠,٠٢٥	١	- /	١٣,٣	١٣,٤٥-١٣,١٥
١,٠٠٠	٤٠			المجموع

الجدول ( ٢ - ٤ )

التكرارات المجمعة

والمجمعة للفئة

التكرار المجمع %	التكرار المجمع	الحدود العليا
٧,٥	٣	$10,45 \geq$
١٧,٥	٧	$10,75 \geq$
٢٧,٥	١١	$11,05 \geq$
٤٥,٥	١٨	$11,35 \geq$
٥٧,٥	٢٣	$11,65 \geq$
٨٠,٠	٣٢	$11,95 \geq$
٩٠,٠	٣٦	$12,25 \geq$
٩٥,٠	٣٨	$12,55 \geq$
٩٥,٠	٣٨	$12,85 \geq$
٩٧,٥	٣٩	$13,15 \geq$
١٠٠,٠	٤٠	$13,45 \geq$

ملاحظات :

١ - س , ترمز إلى مركز الفئة class mark وهو الوسط الحسابي لحدى الفئة ،  
فمثلا مركز الفئة الأولى هو  $\frac{1}{2} (10,45 + 10,15) = 10,3$  .

ويؤخذ مركز الفئة ممثلا لها بمعنى أننا نعتبر أن جميع القيم التي دخلت الفئة  
مساوية لهذا المركز ، فمثلا تضم الفئة الأولى (١٠,٤٥ - ١٠,١٥) ثلاثة من  
الأعداد المعطاة هي ١٠,٤ ، ١٠,٢ ، ١٠,٤ غير أننا في عملية التجميع نلغى هذه  
الأعداد ونعتبر أن بهذه الفئة ثلاثة أعداد كل منها يساوى مركز الفئة وهو ١٠,٣ .

كذلك تضم الفئة الثانية أربعة أعداد هي ١٠,٧ ، ١٠,٥ ، ١٠,٧ ، ١٠,٧ غير أننا نعتبر أن هذه الفئة أربعة أعداد كل منها يساوى مركز الفئة وهو ١٠,٦ . وفي اعتبارنا هذا شيء من التجاوز يسمى بخطأ التجميع ، إلا أن هذه الأخطاء عادة ما يلغى بعضها البعض لأن بعضها بالزيادة والبعض الآخر بالنقصان ، ولاسيما إذا كان حجم التوزيع كبيراً .

٢ - في تكوين الفئات في هذا المثال راعينا أن المتغير هو متغير عددي من النوع المتصل وأن القياس كان إلى أقرب جزء من عشرة من المليمتر . أما إذا اعتبرنا أن القياس مضبوط فيمكن أن نضع الفئات كالآتي :

١٠,٢ - لتعني الفئة التي تشمل الأعداد بدءاً من ١٠,٢ إلى أقل من ١٠,٥  
١٠,٥ - لتعني الفئة التي تشمل الأعداد بدءاً من ١٠,٥ إلى أقل من ١٠,٨  
١٠,٨ - لتعني الفئة التي تشمل الأعداد بدءاً من ١٠,٨ إلى أقل من ١١,١  
وهكذا .. ..

وتستخدم هذه الطريقة أيضاً حين يكون المتغير من النوع الوثناب . وليبان أن هذه الطريقة لا تصلح في الحالة التي تكون فيها البيانات مسجلة بمقياس تقريبي ، اعتبر الحمامة التي سجل طولها على أنه ١٠,٥ ملليمتر ( تقريباً ) . نعلم أن الطول الحقيقي لهذه الحمامة يقع بين العددين ١٠,٤٥ ، ١٠,٥٥ وعلى ذلك فإن الطول الحقيقي قد يكون أصغر من الطول المسجل ١٠,٥ ، مثلاً ١٠,٤٨ ، وفي هذه الحالة ينبغي وضعه في الفئة ١٠,٢ - أو قد يكون أكبر من ١٠,٥ ، مثلاً ١٠,٥٤ ، وفي هذه الحالة ينبغي وضعه في الفئة ١٠,٥ - وما دمنا لا نعرف الطول الحقيقي لهذه الحمامة فإننا نكون في حيرة من استخدام أى من هاتين الفئتين . ونقع في هذه الحيرة أيضاً في تناول كثير من الأطوال الأخرى مثل ١٠,٨ ، ١١,١ ، ١١,٤ ، ٠٠٠ ومن هنا نرى أن هذه الطريقة لا تضمن أن يكون لكل قيمة ( من القيم المقربة ) مكان في واحدة وواحدة فقط من الفئات .

(٢ - ١ - ٣) الجدول التكرارى المزوج ( أو جدول الاقتران ) :

كل من المتالين السابقين يتناول توزيعاً تكرارياً لمتغير واحد ، وفيما يلى مثالان يتناول كل منهما التوزيع التكرارى المشترك لمتغيرين . joint distribution

مثال (٢ - ٣) :

الجدول (٢ - ٥) الآتى يعطى التكرارات المشاهدة لطول محيط الرأس وطول الطفل ساعة الولادة في عينة من ٩٩ مولوداً .

الجدول (٢ - ٥)

المجموع	طول الجسم			محيط الرأس
	٥٣ -	٥٠ -	٤٧ -	
٧٨	٢	٣٦	٤٠	٣٢ -
٢١	٧	١٤	صفر	٣٦ -
٩٩	٩	٥٠	٤٠	المجموع

لدينا متغيران هما ( ١ ) طول محيط الرأس وقد قسمت الأطوال إلى فئتين (٢) طول الجسم وقد قسمت الأطوال إلى ثلاث فئات ، ولهذا يسمى مثل هذا الجدول بجدول اقتران  $3 \times 2$  contingency table لأن المتغيرين يقترنان فيه في توزيع مشترك .

من هذا الجدول نستطيع استخراج الجدولين (٢ - ٦) ، (٢ - ٧) الآتيين :

الجدول (٢ - ٧)  
التوزيع الهامشي لطول الطفل

ك <sub>٢</sub>	طول الجسم
٤٠	- ٤٧
٥٠	- ٥٠
٩	- ٥٣
٩٩	المجموع

الجدول (٢ - ٦)  
التوزيع الهامشي لطول محيط الرأس

ك <sub>٢</sub>	محيط الرأس
٧٨	- ٣٢
٢١	- ٣٦
٩٩	المجموع

يعطى الجدول (٢ - ٦) ما يسمى بالتوزيع الهامشي للمتغير الأول ( طول محيط الرأس ) وهو يعني التوزيع التكرارى لهذا المتغير بصرف النظر عن المتغير الثاني . وبالمثل يعطى الجدول (٢ - ٧) التوزيع الهامشي للمتغير الثاني ( طول الجسم ) .

مثال (٢ - ٤) :

في إحدى التجارب قسم ١٤٦٩ من الرجال في الأعمار ما بين ٦٠ ، ٦٤ عاماً من حيث عادة التدخين إلى قسمين : يدخن ولا يدخن . وبعد ٦ سنوات من بدء التجربة حسب عدد الوفيات للقسمين فنتج التوزيع التكرارى الزدوج المبين بالجدول (٢ - ٨) وهو يعطى التوزيع التكرارى المشترك لمتغيرين من النوع الوصفي هما الوفاة وعادة التدخين .

مثل هذا الجدول يسمى بمجدول اقتران  $2 \times 2$  لأن كلا من المتغيرين مقسم إلى قسمين . استخراج التوزيع الهامشي لكل من المتغيرين .

الجدول (٢ - ٨)

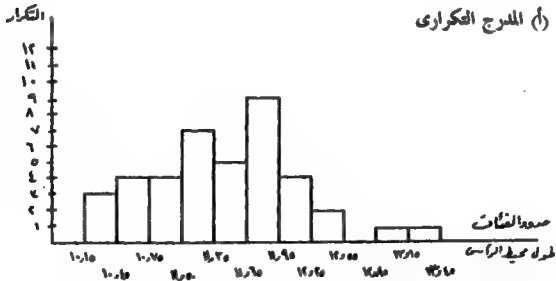
التوزيع للشركات الخاصي الوفلة والتدخين لعينة من كبار السن

الوفلة	التدخين		الاجموع
	يدخن	لا يدخن	
تولي	٥٤	١١٧	١٧١
حي	٣٤٨	٩٥٠	١٢٩٨
الاجموع	٤٠٢	١٠٦٧	١٤٦٩

(٢ - ٢) التمثيل البياني للتوزيعات التكرارية :

المعتاد في تمثيل التوزيعات بيانيا إنشاء محورين متعامدين في المستوى يجرأ كل منهما بمقياس رسم مناسب بحسب الصورة البيانية التي نرغب في تقديمها ، والأشكال الثلاثة الآتية تعرض أشهر هذه الصور وهي تمثل البيانات الواردة بالمثل (٢ - ٢) السابق .

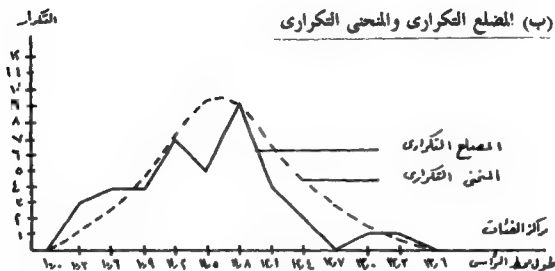
(أ) المدرج التكرارى



الشكل (٢ - ١) المدرج التكرارى لأطوال محطات الرؤوس لعينة من ٤٠ من الحمام المنزل



هذا الشكل يعطى ما يسمى بالمدرج التكرارى ( هستوجرام ) histogram وهو يؤخذ من العمودين الأول والرابع من الجدول (٢ - ٣) ويتألف من عدد من المستطيلات المتلاصقة قواعدها فئات التوزيع وارتفاعاتها تتناسب مع التكرارات المناظرة .

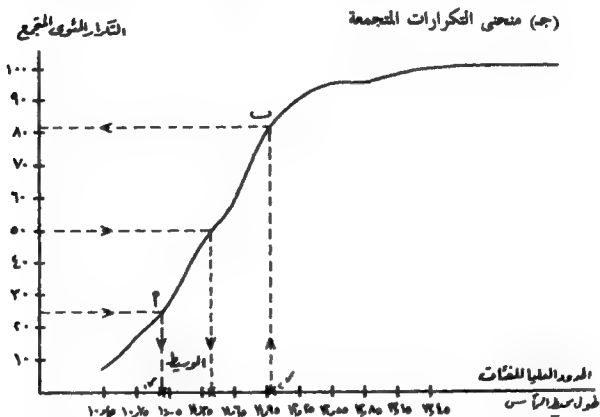


الشكل (٢ - ٢) المضلع التكرارى والمنحنى التكرارى لأطوال عمودات الرؤوس لعدة من ٤٠ من الحمام المنزلى

هذا الشكل يعطى ما يسمى بالمضلع التكرارى frequency polygon وهو يؤخذ من العمودين الثاني والرابع من الجدول (٢ - ٣) . يمثل المحور الأفقى مراكز الفئات ويمثل المحور الرأسى التكرارات ويتبع المضلع من توصيل عدد من النقط ( ١١ نقطة ) إحداثياتها الأفقية مراكز الفئات وإحداثياتها الرأسية التكرارات المناظرة ثم يغلّق المضلع من الطرفين وذلك بتصور وجود فئة إضافية في بداية التوزيع وفئة إضافية في آخره التكرار في كل منهما هو بطبيعة الحال صفر .

كما يعطى هذا الشكل ما يسمى بالمنحنى التكرارى frequency curve وهو منحنى ناعم يمدد باليد ماراً ببعض هذه النقط وقريباً من البعض الآخر ، أى ليس

من الضروري أن يمر بها جميعاً لأن الهدف من رسمه هو محاولة استكشاف الاتجاه العام لتوزيع المتغير في المجتمع الذي أخذت منه العينة ومن الواضح أن عملية التمهيد هذه تعتمد على ذاتية الراسم وتختلف من شخص إلى آخر ، وهي تجرى على أساس أن التوزيع التكرارى الذى لدينا هو توزيع لعينة مأخوذة من مجتمع متصل ، وكلما كبر حجم العينة وصغرت أطوال الفئات كلما اقترب المصطلح التكرارى من المنحنى التكرارى .



الشكل (٢ - ٣) منحنى التكرارات المتجمعة المئوية لأطوال محيطات الرؤوس لعينة من الحمام المنزل .

هذا الشكل يعطى منحنى التكرارات المتجمعة المئوية percentage cumulative frequency curve . وهو يؤخذ من العمودين الأول والثالث من الجدول (٢ - ٤) أى أن الإحداثيات الأفقية للنقط هي الحدود العليا للفئات والإحداثيات الرأسية هي التكرارات المتجمعة المئوية المناظرة . وكان من الممكن أن نرسم المنحنى نفسه من

العمودين الأول والثاني إلا أن هذا يحتاج إلى التفكير في مقياس رسم مناسب لكل توزيع على حدة ، أما استخدام التكرارات المتجمعة المئوية فمن شأنه أن يكون تقسيم المحور الرأسي ثابت لأى توزيع .

ومن هذا المنحنى نستطيع الإجابة إجابة تقريبية عن نوعين من الأسئلة يتمثلان فيما يلى :

( أ ) ما طول محيط رأس الحمامة الذى يقل عنه أو يساويه ٢٥٪ من أطوال محيطات رؤوس الحمام ؟

(ب) ما النسبة المئوية لعدد الحمام الذى تقل أو تساوى أطوال محيطات رؤوسها عن ١٢ ملليمتر ؟

وللإجابة عن السؤال الأول نرسم من النقطة التي تمثل التكرار المتجمع المئوي ٢٥٪ على المحور الرأسي خطا مستقيما يوازي المحور الأفقي ويقطع المنحنى في نقطة ( أ ) ثم نرسم من ( أ ) خطا مستقيما يوازي المحور الرأسي يلقي المحور الأفقي عند النقطة ١٠,٩٨ . وبعملية حسابية بسيطة نجد أن  $10,98 = 1,8$  تقريبا فيكون الطول المطلوب هو ١٠,٩٨ ملليمتر تقريبا . أما الإجابة عن السؤال الثاني فتسير بعكس خطوات الإجابة عن السؤال الأول فنرسم من النقطة التي تمثل العدد ١٢ على المحور الأفقي مستقيما يوازي المحور الرأسي ويقطع المنحنى في نقطة ب ثم نرسم من ب مستقيما يوازي المحور الأفقي ليلقي المحور الرأسي عند النقطة ٨٢ تقريبا فكون النسبة المطلوبة هي ٨٢٪ تقريبا .

ومن منحنى التكرارات المتجمعة المئوية نستطيع بنفس الطريقة أن نوجد تقريبا ما يسمى بالخصائص والريعات وهي أعداد تستخدم في وصف التوزيعات كما سنرى بعد . وهي من المقاييس المسماة بمقاييس الموضع تتميزها عن مقاييس المقدار التي ستقدمها في البند (٢ - ٤) .

إذا كان لدينا توزيع تكرارى لمتغير كى رتبته قيمه تصاعديا فإن المئينات  $١٢$  ،  $٢٢$  ،  $٤٠٠٠$  ،  $٩٩٢$  لهذا التوزيع تعرف بأنها تلك الأعداد التي تقسمه إلى ١٠٠ قسم يشتمل كل منها على عدد متساوى من قيم المتغير أى على ١٪ من هذه القيم . وعلى ذلك فالمئين  $١٢$  هو العدد الذى يقل عنه أو يساويه ١٠٪ من قيم المتغير والمئين  $٨٠$  هو العدد الذى يقل عنه أو يساويه ٨٠٪ من قيم المتغير وهكذا .

وبالمثل الربيعات  $١٢٥$  ،  $٢٥٢$  ،  $٥٠٤$  للتوزيع تعرف بأنها تلك الأعداد التي تقسمه إلى أربعة أقسام يشتمل كل منها على ربع قيم المتغير . ويلاحظ أن :  
الربيع الأول  $١٢٥$  = المئين  $٢٥٢$  = العدد الذى يقل عنه أو يساويه ٢٥٪ من قيم المتغير

$$= ١٠,٩٨ \text{ مليوناً تقريباً في هذا المثال .}$$

الربيع الثاني  $٢٥٢$  = المئين  $٥٠٤$  = العدد الذى يقل عنه أو يساويه ٥٠٪ من قيم المتغير .

$$= ١١,٥ \text{ مليوناً تقريباً في هذا المثال .}$$

ويسمى هذا المقياس أيضاً بالوسيط لأنه يتوسط التوزيع ويقسمه إلى قسمين متساويين في عدد قيم المتغير .  
الربيع الثالث  $٥٠٤$  = المئين  $٧٥٦$  = العدد الذى يقل عنه أو يساويه ٧٥٪ من قيم المتغير .

$$= ١١,٨٨ \text{ مليوناً تقريباً في هذا المثال .}$$

وفي المثال (٢ - ٢) حيث  $٤٠$  حامة نجد أن الربيع الأول وهو ١٠,٩٨ يسبقه عشرة أعداد تقع قيمها من ١٠,٢ إلى ١٠,٩ ، كما نجد أن الوسيط وهو

١١,٥ يسبقه عشرون عددا تقع قيمها من ١٠,٢ إلى ١١,٥ ، ونجد أن الربيع الثالث وهو ١١,٨٨ يسبقه ثلاثون عددا تقع قيمها من ١٠,٢ إلى ١١,٨ .  
ونستطيع إيجاد المتينات والربيعات بطريقة حسابية وهي طريقة أكثر دقة نوضحها عن طريق التوزيع الذي بالمثل (٢ - ٢) والذي يمكن تلخيصه بالجدول (٢ - ٩) الآتي :

#### الجدول (٢ - ٩)

التوزيع التكرارى والتوزيع للجمع لأطوال عمقات الرؤوس بالمحطات لجهة من المحطات لتقلى

التكرار المتجمع	التكرار	الفئة
٣	٣	١٠,٤٥ - ١٠,١٥
٧	٤	١٠,٧٥ - ١٠,٤٥
١١	٤	١١,٠٥ - ١٠,٧٥
١٨	٧	١١,٣٥ - ١١,٠٥
٢٣	٥	١١,٦٥ - ١١,٣٥
٣٢	٩	١١,٩٥ - ١١,٦٥
٣٦	٤	١٢,٢٥ - ١١,٩٥
٣٨	٢	١٢,٥٥ - ١٢,٢٥
٣٨	٠	١٢,٨٥ - ١٢,٥٥
٣٩	١	١٣,١٥ - ١٢,٨٥
٤٠	١	١٣,٤٥ - ١٣,١٥

لايجاد الوسيط  $\frac{3}{4}$  والريعين  $\frac{1}{4}$  ،  $\frac{3}{4}$  :

ترتيب الوسيط  $= \frac{3}{4}$  ( هذه قاعدة عامة لتوزيعات المتغيرات المتصلة )

$$\frac{3}{4} = \frac{20}{20} = 20$$

إذن الوسيط هو الطول الذى يقل عنه أو يساويه أطوال ٢٠ حمامة .

نلاحظ من الجدول أن هناك ١٨ حمامة تقل أطوالها عن ١١,٣٥ مليلمترا .

وأن هناك ٢٣ حمامة تقل أطوالها عن ١١,٦٥ مليلمترا .

وعلى ذلك يجب أن يقع الوسيط في الفئة ١١,٣٥ - ١١,٦٥ وتسمى هذه الفئة حيثثذ

بالفئة الوسيطة . إن هذه الفئة تشتمل على ٥ وحدات ( حمامات ) نريد أن نأخذ منها

اثنين فقط لنستكمل العدد المطلوب ( من ١٨ حمامة إلى ٢٠ حمامة ) وعلى فرض أن

الوحدات الخمسة موزعة بانتظام داخل الفئة بمعنى أنها تبعد عن بعضها بمساافات

متساوية فإن العدد ٢ يمثل طولاً قدره  $\frac{2}{5}$  طول الفئة أى طولاً قدره  $\frac{2}{5} \times 0,3 = 0,12$

٠,١٢ مم ولما كانت بداية الفئة ١١,٣٥ مم فإن :

$$\text{الوسيط} = \frac{3}{4} = 11,35 + 0,12 = 11,47 \text{ مليلمترا .}$$

وبصفة عامة نوجد الوسيط من الصيغة الآتية :

الوسيط = الحد الأدنى للفئة الوسيطة +

ترتيب الوسيط - التكرار المتجمع للفئة السابقة للفئة الوسيطة  $\times$  طول الفئة

التكرار في الفئة الوسيطة

وبنفس المنطق السابق نوجد الريعين الأول والثالث بالصيغتين الآتيتين مع

ملاحظة أن ترتيب الريع الأول هو  $\frac{1}{4}$  وترتيب الريع الثالث هو  $\frac{3}{4}$  وذلك

بالنسبة للتوزيعات التكرارية للمتغيرات المتصلة .

$$s_1 = \text{الحل الأول} + \text{الحد الأدنى لفئة الربيع الأول} +$$

$$\text{ترتيب } s_1 - \frac{\text{التكرار المتجمع للفئة السابقة لفئة الربيع الأول} \times \text{طول الفئة}}{\text{التكرار في فئة الربيع الأول}}$$

$$s_2 = \text{الحل الأول} + \text{الحد الأدنى لفئة الربيع الثالث} +$$

$$\text{ترتيب } s_2 - \frac{\text{التكرار المتجمع للفئة السابقة لفئة الربيع الثالث} \times \text{طول الفئة}}{\text{التكرار في فئة الربيع الثالث}}$$

في المثال (٢ - ٢) نجد ما يلي :

$$\text{ترتيب الربيع الأول} = \frac{1}{4} = 0,25$$

$$s_1 = 0,25 \times \frac{10-0}{4} + 10,75 = 0,3 \times 2,5 + 10,75 = 0,75 + 10,75 = 11,5$$

= ١٠,٩٧٥ مليلترا .

$$s_2 = \frac{1}{4} \times \frac{10-0}{4} = 0,625$$

$$s_2 = 0,625 \times \frac{11-0}{4} + 11,65 = 0,3 \times 2,8125 + 11,65 = 0,84375 + 11,65 = 12,49375$$

= ١١,٨٨٣ مليلترا .

وتطبق هذه الصيغ أيضا في حالة التوزيعات التكرارية البسيطة وفي حالة التوزيعات غير التكرارية أي التي على صورة مجموعة من القيم  $s_1, s_2, \dots$  ،  $s_r$  ( تكرار كل منها الواحد الصحيح ) ، بشرط أن يكون المتغير من النوع المتصل فتجبر أي قيمة  $s$  من قيم المتغير كأنها فترة تبدأ بنصف وحدة خطوة أسفل العدد  $s$  وتنتهي بنصف وحدة خطوة أعلاه وبذلك يكون طول الفترة مساويا الواحد إذا كانت  $s$  مقدرية بالوحدات الصحيحة ويساوى ٠,١ إذا كانت مقدرية إلى خانة عشرية واحدة ويساوى ٠,٠١ إذا كانت مقدرية إلى خاتين عشريتين وهكذا . ونوضح ذلك بأخذ المثال (٢ - ١) الذي تلخصه بالجدول (٢ - ١٠) الآتي :

المجدول (٢-١٠)

س	ك	س
٢	٢	٧٧
٣	١	٧٨
٥	٢	٧٩
٧	٢	٨٠
٨	١	٨١
١٠	٢	٨٢
١٢	٢	٨٣
١٤	٢	٨٤
١٧	٣	٨٥
٢٢	٥	٨٦
٢٥	٣	٨٧

$$\text{ترتيب الربيع الأول} = \frac{٢٥}{٤} = ٦,٢٥$$

الربيع الأول يقع في الفئة التي تعبر عن العدد ٨٠ التي تمتد من ٧٩,٥ إلى ٨٠,٥ والتكرار فيها ٢

$$٨٠,١٢٥ = ١ \times \frac{٥-٦,٢٥}{٤} + ٧٩,٥ = ١,$$

$$١٢,٥ = \frac{٢٥}{٢} = \text{ترتيب الوسيط}$$

الوسيط يقع في الفئة التي تعبر عن العدد ٨٤ التي تمتد من ٨٣,٥ إلى ٨٤,٥ والتكرار فيها ٢

$$٨٣,٧٥ = ١ \times \frac{١٢-١٢,٥}{٢} + ٨٣,٥ = ٢,$$

$$١٨,٧٥ = \frac{٢٥ \times ٢}{٤} = \text{ترتيب الربيع الثالث}$$

$$٨٥,٨٥ = ١ \times \frac{١٨-١٨,٧٥}{٥} + ٨٥,٥ = ٣,$$



**ملاحظة :** لا تصلح هذه الطريقة ولا طريقة المنحنى المتجمع في حالة توزيعات المتغيرات غير المتصلة ( الوثابة ) ، على أنه في حالة التوزيعات غير التكرارية من الواضح أن الوسيط هو القيمة التي تقع في وسط التوزيع إذا كان عدد قيم المتغير فردياً أو هو متوسط القيمتين الوسطيتين إذا كان عدد القيم زوجياً . فمثلاً للمجموعة .

٤ ٥ ٥ ٦ ٧ ٧ ٩ ١١ ١٢ ١٢ ١٤ ١٥ ١٥ ٤٠ ٥٠

التي عدد قيمها ١٥ يكون الوسيط هو العدد ١١ إذ أن هذا العدد يقسم المجموعة إلى قسمين متساويين في عدد القيم سبعة منها قبله وسبعة منها بعده .

أما بالنسبة للمجموعة

٤ ٥ ٥ ٦ ٧ ٧ ٩ ١١ ١٢ ١٢ ١٤ ١٥ ١٥ ٤٠

التي عدد قيمها ١٤ فيكون الوسيط هو العدد  $\frac{11+12}{2} = 11.5$  إذ أنه يقسم المجموعة إلى قسمين متساويين في عدد القيم سبعة منها قبله وسبعة منها بعده مع ملاحظة أن الوسيط هنا ليس أحد قيم المجموعة المعطاة .

بصفة عامة إذا رتب قائمة من الأعداد حجمها  $n$  لمتغير ولاب ترتيباً تصاعدياً فإن :

$$(أ) \text{ ترتيب الوسيط هو } \frac{1}{2} (n + 1)$$

ففي القائمة الأولى لدينا  $n = 15$  . ترتيب الوسيط ٨ وعلى ذلك فالوسيط هو العدد الثامن في القائمة أي العدد ١١ .

وفي القائمة الثانية لدينا  $n = 14$  . ترتيب الوسيط ٧,٥ وهذا العدد يعنى أن الوسيط هو متوسط العددين السابع والثامن في القائمة أي العدد ١٠ .

(ب) بحسب ترتيب الربيع الأول من ترتيب الوسيط كالآتي :

ترتيب  $\frac{1}{4}n$  هو  $\frac{1}{4}$  (ترتيب الوسيط بعد حذف الكسر إذا وجد  $n + 1$ ) .

(ج) من المائل يكون ترتيب الربع الثالث هو نفس ترتيب الربع الأول حين نقرأ قائمة الأعداد عكسياً من النهاية إلى البداية .

ففي القائمة الأولى لدينا  $n = 10$  ، ترتيب الوسيط هو 8

$$\therefore \text{ترتيب } M_1 \text{ هو } \frac{1}{2}(1 + 8) = 4,5$$

وهذا يعنى أن الربع الأول هو متوسط العددين الرابع والخامس في القائمة أى أن

$$M_1 = \frac{1}{2}(7+6) = 6,5$$

$$M_2 = \frac{1}{2}(14+10) = 12,5 \quad \text{من المائل}$$

وفي المجموعة الثانية لدينا  $n = 14$  ، ترتيب الوسيط هو 7,5

$$\therefore \text{ترتيب } M \text{ هو } \frac{1}{2}(1+7) = 4$$

$$M_1 = 6$$

$$M_2 = 16 \quad \text{من المائل}$$

## ( ٢ - ٤ ) الوصف العددي للتوزيعات التكرارية :

حين يكون لدينا توزيع لتغير عددي وحين يكون لهذا التوزيع قمة واحدة كما يبدو مثلاً من المنحني التكراري ، فإننا نستطيع وصفه موضوعياً من حيث عدة جوانب أهمها :

**Central Tendency**

**Dispersion**

**Skewness**

**Kurtosis**

( أ ) النزعة المركزية

(ب) التشتت

(ج) الإلتواء

( د ) التفرطح

ففي أغلب التوزيعات ذات القمة الواحدة يتراكم عدد كبير من قيم المتغير حول قيمة معينة ويقل هذا التراكم تدريجياً على وجه العموم كلما ابتعدت القيم عن هذه القيمة من الجانبين . هذا التراكم أو التجمع حول قيمة ما يسمى بالنزعة المركزية للتوزيع أو بالقيمة المتوسطة وتسمى القيمة التي يحدث حولها التجمع بمقياس النزعة المركزية . وبهنا في دراسة التوزيعات الحصول على هذا المقياس . ولما كان هذا المقياس يختلف في تركيبه بحسب طبيعة البيانات والهدف من دراستها فقد وضعت عدة مقاييس للنزعة المركزية نختار منها ما نرى أنه يلائم ما بأيدينا من توزيعات . ومن أشهر هذه المقاييس ما يلي :

الوسط الحسابي - الوسط - المنوال - الوسط الهندسي - الوسط التوافقي .

وسنهم هنا بصفة خاصة بالوسط الحسابي لأسباب عدة منها أنه أقوى مقاييس النزعة المركزية استجابة للمعالجة الرياضية ، وهو المقياس الذي نستخدمه عادة ما لم يتضح لنا أنه لا يعبر تعبيراً صادقاً عن هذه النزعة كما هو الحال مثلاً عندما يكون التوزيع مشتملاً على قيم متطرفة تشذ عن بقية القيم .

كما يهنا كذلك قياس تشتت التوزيع أي قياس مدى انحراف قيم المتغير بالنسبة إلى بعضها وبالنسبة إلى القيمة المتوسطة ، أو بمعنى معكوس ، مدى تجانس التوزيع . ومن أشهر مقاييس التشتت ما يلي :

المدى - الانحراف الربيعي - الانحراف المتوسط - الانحراف المعياري - معامل الاختلاف .

وسنهم هنا بصفة خاصة بالانحراف المعياري لأنه كمقياس للتشتت يتمشى مع الوسط الحسابي كمقياس للنزعة المركزية وهما يؤخذان معاً أو يتركان معاً .

## ( ٢-٤-١ ) الوسط الحسابي والانحراف المعياري :

### MEAN AND STANDARD DEVIATION

مثال ( ٢-٥ ) :

اعتبر الأعداد السبعة الآتية : ٥ ٨ ٧ ١٠ ١١ ١٤ ٩

نعلم أن الوسط الحسابي لعدد من القيم هو مجموع هذه القيم مقسوماً على عددها وهذا تعريف عام للوسط الحسابي ويمكن صياغته رمزياً في حالتنا هذه كالآتي .  
(١)

$$\text{الوسط الحسابي} = \bar{x} = \frac{\text{مجموع القيم}}{n}$$

حيث  $\bar{x}$  تعبر عن قيم المتغير ،  $n$  تعبر عن عدد هذه القيم .

ويعرف التباين Variance بأنه متوسط مجموع مربعات انحرافات قيم المتغير عن الوسط الحسابي وهذا ما نستطيع كتابته رمزياً كالآتي ( على أساس أن التوزيع التكراري هو توزيع عينة ) :

$$\text{التباين} = s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \quad (٢-١)$$

يلاحظ أن هذا العدد يساوي صفراً إذا وإذا فقط تساوت جميع القيم  $\bar{x}$  لأن كلا منها في هذه الحالة يساوي  $\bar{x}$  ، وهناك صورة أخرى للتباين تشتق من هذه الصورة بعمليات جبرية بسيطة ، وهذه الصورة هي :

$$s^2 = \frac{1}{n-1} \left[ \sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] \quad (٢-ب)$$

وبالرغم من أن هاتين الصورتين متطابقتان رياضياً ، إلا أننا في حساب التباين نستخدم الصورة الثانية لأن الخطأ الذي ينتج فيها من تقريب الأعداد أقل من ذلك

الذي ينتج من الصورة الأولى ، كما أنها أكثر طواعية لحاسبات الجيب والحاسبات الإلكترونية .

أما الانحراف المعياري فيعرف بأنه الجذر التربيعي الموجب للتباين ويرمز له بالرمز  $\sigma$  . وهو عدد موجب دائماً بالتعريف .

من الصيغتين ( ١ ) ، ( ٢ - ب ) نرى أن حساب الوسط الحسابي والتباين يعتمد على حساب المجموعين  $\sum x$  ،  $\sum x^2$  ، وهذان المجموعان يمكن إيجادهما مباشرة من حاسبات الجيب أو من الجدول ( ٢ - ١١ ) الآتي وهو يخص المثال ( ٢ - ٥ ) .

الجدول ( ٢ - ١١ )

لإيجاد الوسط الحسابي والانحراف المعياري لمجموعة من القيم

س <sub>ر</sub>	س <sub>٢</sub>
٥	٢٥
٨	٦٤
٧	٤٩
١٠	١٠٠
١١	١٢١
٤	١٦
٩	٨١
٥٤	٤٥٦

$$\text{الوسط الحسابي } \bar{X} = \frac{\text{مجموع}}{n} = \frac{54}{7} = 7,71 \text{ تقريباً}$$

$$\text{التباين } s^2 = \frac{1}{n-1} [\text{مجموع } s^2 - \frac{(\sum X)^2}{n}] = \frac{1}{1-7} [\frac{(54)^2}{7} - 406] = \frac{1}{1-7} [432,86 - 406] = 3,95$$

$$= 6,071 \text{ تقريباً}$$

$$\text{الانحراف المعياري } s = \sqrt{6,071} = 2,46 \text{ تقريباً}$$

مثال ( ٢ - ٦ ) :

أوجد الوسط الحسابي والانحراف المعياري للتوزيع التكراري البسيط المعطى بالمثل ( ٢ - ١ ) السابق .

نظراً لأن كل قيمة  $s_r$  من قيم المتغير مكررة  $K_r$  من المرات فإن مجموع القيم يكون  $s_r K_r$  وعلى ذلك فإن التعريف العام للوسط الحسابي يمكن صياغته في هذه الحالة كالآتي :

$$(٣) \quad \text{الوسط الحسابي } \bar{X} = \frac{\sum s_r K_r}{n}$$

$$\text{حيث } n = \sum K_r$$

وبالمثل نعرف التباين كالآتي :

$$(٤ - أ) \quad \text{التباين } s^2 = \frac{1}{n-1} [\sum s_r^2 K_r - \frac{(\sum s_r K_r)^2}{n}]$$

$$\text{أو } (٤ - ب) \quad s^2 = \frac{1}{n-1} [\sum s_r^2 K_r - \frac{(\sum s_r K_r)^2}{n}]$$

يلاحظ أن التعريفين ( ١ ) ، ( ٢ ) ما هما إلا حالة خاصة من التعريفين ( ٣ ) ، ( ٤ ) تكون فيها جميع التكرارات  $K_r$  مساوية للواحد .  
 لإيجاد المجموعين  $\sum K_r$  ،  $\sum K_r^2$  نوظف لحساب الوسط الحسابي والتباين نستخدم جدولاً كالجدول ( ٢ - ١٢ ) الآتي :

الجدول ( ٢ - ١٢ )

إيجاد الوسط الحسابي والانحراف المعياري لموزع تكراري بسيط

$\sum K_r$	$\sum K_r^2$	$K_r$	$K_r$
١١٨٥٨	١٥٤	٢	٧٧
٦٠٨٤	٧٨	١	٧٨
١٢٤٨٢	١٥٨	٢	٧٩
١٢٨٠٠	١٦٠	٢	٨٠
٦٥٦١	٨١	١	٨١
١٣٤٤٨	١٦٤	٢	٨٢
١٣٧٧٨	١٦٦	٢	٨٣
١٤١١٢	١٦٨	٢	٨٤
٢١٦٧٥	٢٥٥	٣	٨٥
٣٦٩٨٠	٤٣٠	٥	٨٦
٢٢٧٠٧	٢٦١	٣	٨٧
١٧٢٤٨٥	٢٠٧٥	٢٥	

$$\text{الوسط الحسابي} = \bar{x} = \frac{\text{مجموع القيم}}{n} = \frac{2070}{20} = 103$$

$$\text{التباين} = s^2 = \frac{1}{n-1} [\text{مجموع القيم}^2 - \frac{(\text{مجموع القيم})^2}{n}]$$

$$= \frac{1}{19} [2070^2 - \frac{(2070)^2}{20}] = 10,82$$

$$\text{الانحراف المعياري} = s = 3,291$$

مثال ( ٢ - ٧ ) :

أوجد الوسط الحسابي والانحراف المعياري للتوزيع التكراري ذي الفئات المعطى  
بالمثال ( ٢ - ٢ ) السابق .

في حالة التوزيع التكراري ذي الفئات نعتبر كما سبق الذكر أن جميع القيم الواقعة في فئة ما مساوية لمركز الفئة . وعلى ذلك يكون مجموع هذه القيم مجموع تكرار مركز فئة حيث مركز الفئة ، وتكون التعاريف ( ٣ ) ، ( ٤ - أ ) ، ( ٤ - ب ) صالحة للاستخدام هنا لإيجاد الوسط الحسابي والتباين مع فارق واحد وهو أن مركز فئة ترمز إلى مراكز الفئات في حالة التوزيعات التكرارية ذات الفئات ، بينما ترمز إلى قيم المتغير في حالة التوزيعات التكرارية البسيطة . ويجري الحساب كما في الجدول ( ٢ - ١٣ ) الآتي :



الجدول (٢-١٣)

إيجاد الوسط الحسابي والانحراف المعياري لتوزيع تكراري ذي فئات

الفئات	م ر	ك ر	ك ر م ر	ك ر م ر
١٠,١٥ - ١٠,٤٥	١٠,٣	٣	٣٠,٩	٣١٨,٢٧
١٠,٤٥ - ١٠,٧٥	١٠,٦	٤	٤٢,٤	٤٤٩,٤٤
١٠,٧٥ - ١١,٠٥	١٠,٩	٤	٤٣,٦	٤٧٥,٢٤
١١,٠٥ - ١١,٣٥	١١,٢	٧	٧٨,٤	٨٧٨,٠٨
١١,٣٥ - ١١,٦٥	١١,٥	٥	٥٧,٥	٦٦١,٢٥
١١,٦٥ - ١١,٩٥	١١,٨	٩	١٠٦,٢	١٢٥٣,١٦
١١,٩٥ - ١٢,٢٥	١٢,١	٤	٤٨,٤	٥٨٥,٦٤
١٢,٢٥ - ١٢,٥٥	١٢,٤	٢	٢٤,٨	٣٠٧,٥٢
١٢,٥٥ - ١٢,٨٥	١٢,٧	٠	٠	٠
١٢,٨٥ - ١٣,١٥	١٣,٠	١	١٣,٠	١٦٩,٠٠
١٣,١٥ - ١٣,٤٥	١٣,٣	١	١٣,٣	١٧٦,٨٩
المجموع		٤٠	٤٥٨,٥	٥٢٧٤,٤٩

الوسط الحسابي لمخطط رأس الحمامة =  $\bar{X} = \frac{458.5}{40} = 11.4625$  مليمترا

$$\frac{[ \frac{(458.5)}{40} - 5274.49 ] \cdot \frac{1}{39}}{40} = \sigma^2 =$$

$$0.4800 =$$

الانحراف المعياري =  $\sqrt{0.4800} = 0.6928$  مليمترا

### ملاحظات :

( ١ ) الأصل في تعريف التباين هو  $\frac{1}{n} \sum (s_i - \bar{s})^2$  ولكن حين نستخدم تباين عينة حجمها  $n$  ووسطها الحسابي  $\bar{s}$  لتقدير تباين مجتمع تقديراً غير متحيز نضع  $\frac{1}{n-1}$  بدلاً من  $\frac{1}{n}$  لأن تباين العينة يكون عادة أصغر من تباين المجتمع وهذا التعديل يجعل تباين العينة أكثر ملاءمة لتقدير تباين المجتمع ، وهناك من النظريات الرياضية ما يؤيد ذلك . وفيما عدا هذه الحالة نستخدم الصيغة الأصلية للتباين أي نحفظ بالعدد  $n$  .

( ٢ ) لا تتغير قيمة التباين إذا طرح ( أو أضيف ) أي عدد من جميع قيم المتغير .

### ( ٢-٤-٢ ) معامل الاختلاف :

يعرف معامل الاختلاف  $m$  .  $x$  . لتوزيع ما كآتي :

$$m \cdot x = \frac{s}{\bar{s}} \times 100 \quad (٥)$$

حيث  $s$  هو الانحراف المعياري للتوزيع ،  $\bar{s}$  وسطه الحسابي .

فمثلاً ، للتوزيع الذي بالمثال ( ٢-٦ ) :

$$m \cdot x = \frac{3.965}{83} \times 100 = 47.77$$

وللتوزيع الذي بالمثال ( ٢-٧ ) :

$$m \cdot x = \frac{6.079}{11.4625} \times 100 = 52.99$$

إن معامل الاختلاف هو مقياس مطلق للتشتت وهو يستخدم لمقارنة تشتتات التوزيعات خاصة في الحالتين الآتيتين :

( أ ) حين تختلف متوسطات التوزيعات اختلافاً كبيراً كما هو الحال عند مقارنة تشتت أطوال أذبال الأفيال وتشتت أطوال أذبال الفيران .  
 ( ب ) حين تختلف الوحدات التي تقاس بها المتغيرات كما هو الحال عند مقارنة تشتت أطوال مجتمع ما بأوزان هذا المجتمع .

### SKENNESS

( ٢ - ٤ - ٣ ) الالتواء :

التواء توزيع ما يعني مدى بعده عن التمثال . ويكون التوزيع التكراري متاثلاً إذا كانت التكرارات موزعة توزيعاً متاثلاً حول الوسط الحسابي ، بمعنى أن تكون لقيم المتغير المتساوية البعد عن الوسط الحسابي نفس التكرارات . والتوزيع الآتي مثال لذلك :

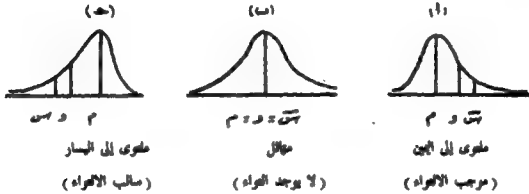
س	١	٢	٤	٦	٨	١٠	١٢	١٤	١٥
ك	٣	١٠	٣٥	٨٠	٩٥	٨٠	٣٥	١٠	٣

فهذا توزيع متاثلاً حول وسطه الحسابي  $\bar{x} = 8$

ويبدو التمثال جلياً من النظر إلى المضلع التكراري المين بالشكل ( ٢ - ٤ ) ( التكرار ) والذي يمثل هذا التوزيع .



وتقسم المنحنيات التكرارية من حيث الالتواء إلى ثلاثة أنواع تبين من الشكل ( ٢ - ٥ ) الآتي حيث  $\bar{x}$  ترمز إلى الوسط الحسابي ، و  $\sigma$  ترمز إلى الوسيط ،  $\mu$  ترمز إلى المتوسط وهو القيمة الأكثر تكرارا في التوزيع .



الشكل ( ٢ - ٥ ) : أنواع الالتواء

إذا كانت هذه الأشكال تمثل توزيعا لدرجات طلاب في امتحان ما فالشكل ( أ ) يشير إلى أن عددا كبيرا من الطلاب حصلوا على درجات أقل من المتوسط مما قد يعني أن مستوى الطلاب أقل من مستوى الامتحان . . . ، والشكل ( ج ) يشير إلى عكس ذلك . وهناك حقيقة هامة مثلت بوضوح في هذه الأشكال نقدمها كما يلي :

إذا كان التوزيع ملتويا إلى اليمين فإن  $\bar{x} < \sigma < \mu$  والعكس بالعكس .  
 وإذا كان التوزيع متائلا فإن  $\bar{x} = \sigma = \mu$  والعكس بالعكس .  
 وإذا كان التوزيع ملتويا إلى اليسار فإن  $\bar{x} > \sigma > \mu$  والعكس بالعكس .  
 ويقاس الالتواء عادة بأحد المقياسين الآتين :

$$(١) \text{ مقياس الالتواء } = \frac{3(\text{الوسط الحسابي} - \text{الوسيط})}{\text{الانحراف المعياري}} = \frac{3(\bar{x} - \sigma)}{\sigma}$$

وهذا للمقياس مبنى على الحقيقة سابقة الذكر .



إن وصف المنحنيات بأنها مدببة أو مفرطحة يكون بالمقارنة مع المنحنيات المعتدلة التي سنتناولها بالدراسة في فصل قادم . وحين نقول إن المنحني مدبب فنحن نعني أن عدداً كبيراً من المفردات يتراكم بالقرب من الوسط الحسابي وعند الذيلين ولا يكون بالمواضع الأخرى إلا عدداً قليلاً منها ، وذلك بالمقارنة مع المنحنيات المعتدلة . كذلك حين نقول إن المنحني مفرطح فنحن نعني أن عدداً قليلاً من المفردات يتراكم بالقرب من الوسط الحسابي وعند الذيلين ويكون هناك عدد كبير منها بالمواقع الأخرى ، وذلك بالمقارنة مع المنحنيات المعتدلة .

ويعرف المقياس الذي سنأخذه للتفرطح كالآتي وهو يتمشي مع الوسط الحسابي والتباين ومعامل الالتواء السابق تعريفها وجميعها من فصيلة تسمى بفصيلة العزوم :

$$\text{معامل التفرطح} = \frac{1}{n-1} \sum r (s_r - \bar{s})^2 / \sigma^2 \quad (8)$$

حيث  $\sigma$  ترمز إلى الانحراف المعياري . وإذا وجد أن قيمة هذا المعامل في عينة ما قريبة من العدد 3 قيل إن المنحني معتدل التفرطح ، وإذا زادت عن هذا العدد قيل إن المنحني مدبب ، وإذا قلت قيل إنه مفرطح .

ملاحظة :

إن الوسط الحسابي والانحراف المعياري ومعامل الالتواء ومعامل التفرطح ، بالإضافة إلى حجم التوزيع هي جل ما نحتاج إليه في التحليل الوصفي للتوزيعات ذات القمة الواحدة ، وإذا كان التوزيع يمثل عينة لمجتمع ما فإن هذه القيم تتخذ أساساً لتقدير المعالم الإحصائية لهذا المجتمع باستخدام الطرق الإحصائية كما سنرى بعد . على أن هناك توزيعات لا تصلح هذه المقاييس لوصفها ويستلزم الأمر حينئذ اختيار مقاييس أخرى تناسب هذه التوزيعات . فمثلاً حين يكون التوزيع شديد الالتواء أو محوياً على قيم متطرفة تشذ عن بقية القيم لا يكون الوسط الحسابي معبراً تعبيراً صادقاً عن النزعة المركزية ولا يكون الانحراف المعياري معبراً تعبيراً صادقاً عن التشتت ويوضح هذا من المثال الآتي :

## مقال ( ٢ - ٨ ) :

القيم الآتية هي أعمار ١٥ مريضاً بالسنوات دخلوا أحد أقسام إحدى المستشفيات في يوم ما ، وذلك بعد ترتيب هذه القيم تصاعدياً :

٦٠ ٥٠ ١٤ ١٣ ١٢ ١١ ١١ ١٠ ٨ ٧ ٦ ٥ ٤ ٣ ٢

نلاحظ أن هناك قيمتين تشذان عن بقية القيم وهما ٦٠ ، ٥٠ وإذا حسبنا الوسط الحسابي لهذه المجموعة نجده يساوي  $\frac{217}{15} = 14,5$  سنة ولا يعقل

أخذ هذه القيمة للتعبير عن متوسط أعمار المرضى فهي تزيد عن جميع قيم المجموعة المعطاة ماعدا قيمتين وفي الوقت ذاته تقل كثيراً عن هاتين القيمتين .

ويفضل في هذه الحال استخدام الوسيط كمقياس للترعة المركزية لأنه لا يتأثر بالقيم المتطرفة . والوسيط هنا هو العدد ١٠ ومن الواضح أن هذا العدد يتوسط التوزيع وهو أصدق تعبيراً عن متوسط الأعمار من الوسط الحسابي .

في مثل هذه الحالات يستخدم مايسمى بنصف المدى الربيعي لقياس التشتت ومايسمى بمعامل الالتواء الربيعي لقياس الالتواء وهما مقياسان يتمشيان مع الوسيط ويعرفان بدلالة الربيعات كالآتي :

نصف المدى الربيعي  $= \frac{1}{4} (P_3 - P_1)$  (٩) Semi-interquartile range

معامل الالتواء الربيعي  $= \frac{(P_3 - P_1) - (P_2 - P_1)}{(P_3 - P_1) + (P_2 - P_1)}$  (١٠)

$$= \frac{P_3 + P_2 - 2P_1}{P_3 - P_1}$$

وبالطبي يمكن استخدامه للمقارنة بين التواءات التوزيعات . وهو يساوي صفرأ للتوزيعات المتأثلة حيث يقع الربيان الأول والثالث على بعدين متساوين من الوسيط م . كما يلاحظ أن قيمة هذا المقياس تقع بين العددين - ١ ، + ١ وكلما ابتعدت قيمته عن الصفر من اليمين أو اليسار كلما دل ذلك على التواء التوزيع . كذلك :

$$(١١) \quad \text{معامل الاختلاف الربي} = \frac{\text{نصف المدى الربي}}{\text{الوسيط}} \times ١٠٠$$

وللمثال ( ٢ - ٨ ) الأخير نجد - كما في البند ( ٢ - ٣ - ١ ) ومع ملاحظة أن المتغير متصل - مايلي :

$$١٢,٧٥ = م , ١٠ = م , ٥,٢٥ = م$$

$$\text{نصف المدى الربي} = \frac{١}{٢} (٥,٢٥ - ١٢,٧٥) = ٣,٧٥$$

$$\text{معامل الالتواء الربي} = \frac{٥,٢٥ + ٢٠ - ١٢,٧٥}{٥,٢٥ - ١٢,٧٥} = ٠,٢٧$$

## تمارين ( ٢ - ١ )

( ١ ) احسب الوسط الحسابي والانحراف المعياري ومعامل الاختلاف للأعداد الآتية

$$١٠٠ \quad ٩٨ \quad ١٢٠ \quad ٩٥ \quad ١١٥ \quad ١١٠ \quad ١٠٠ \quad ١٢٥ \quad ٩٢ \quad ١٣٠$$



( ٢ ) رصدت أعمار عينة من ٢٧ شخصا بالسنوات المختلفة عند إصابتهم بمرض ما فوجدت كالآتي :

٣٩ ٥٠ ٢٦ ٤٥ ٤٧ ٧١ ٥١ ٣٣ ٤٠ ٤٠ ٥١ ٥٩ ٤٨ ٤٤ ٤٧ ٦١  
٤١ ٥٧ ٣٦ ٥٥ ٦٣ ٦٦ ٤٢ ٥٤ ٤٧ ٥٣ ٥٤

أوجد  $\bar{x}$  ،  $s^2$  ، ومنها احسب الوسط الحسابي والانحراف المعياري للعمر عند الإصابة بذلك المرض . ( لا داعي لتكوين توزيع تكراري ) .

( ٣ ) فحص ١٢٢ قرنا من قرون شجرة السرقم ( الأباتوس الكاذب *laburnum* ) فوجد مايلي :

عدد البنور في القرن	١	٢	٣	٤	٥	٦	٧	٨
عدد القرون	٢٧	٣٢	٢١	١٥	١٣	٩	٤	١

( أ ) احسب الوسط الحسابي والانحراف المعياري لعدد البنور في القرن .

( ب ) ارسم المدرج التكراري للتوزيع .

( ٤ ) قيس أطوال ٢٥ عظمة فخذ نوع من الحشرات ( م م ١٠× )<sup>١</sup> فوجدت كما يلي :

٤,٤ ٣,٩ ٣,٨ ٣,٩ ٤,٢ ٣,٩ ٤,٣ ٤,٣ ٣,٣ ٤,٣ ٣,٥ ٤,٣ ٣,٦ ٣,٨  
٤,١ ٤,٤ ٣,٦ ٤,٥ ٤,٤ ٣,٦ ٤,١ ٣,٦ ٤,٧ ٣,٨

( أولا ) كون جدولا تكراريا ذات فئات طول فته ٠,٣٠ ومثله بيانيا بمضلع تكراري .

( ثانيا ) احسب كلا من الوسط الحسابي  $\bar{x}$  والانحراف المعياري  $s$  لطول عظمة الفخذ .

( ثالثاً ) ارسم منحني التكرارات المتجمعة المتوية ومنه احسب مايلي :

( أ ) النسبة المتوية لعدد الحشرات التي تقل أطوالها عن ٤ م م<sup>١٠٠</sup>

( ب ) النسبة المتوية لعدد الحشرات التي تقع أطوالها بين العددين ٥±٣

( ج ) الوسيط أي طول عظمة الفخذ الذي تقل عنه أطوال ٥٠٪ من الحشرات .

( ٥ ) التوزيعات الثلاثة الآتية هي توزيعات درجات مجموعة من ٢٤ طالبا في ثلاثة اختبارات . ارسم المصّلع التكراري لكل منها واذكر تعليقاً عن التواء كل توزيع .

التوزيع الأول س<sub>ر</sub> : ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩ ١٠

ك<sub>ر</sub> : ٤ ٦ ٥ ٢ ٢ ١ ١ ١ ١ ١

التوزيع الثاني س<sub>ر</sub> : ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩ ١٠

ك<sub>ر</sub> : ١ ١ ٢ ٣ ٥ ٥ ٣ ٢ ١ ١

التوزيع الثالث س<sub>ر</sub> : ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩ ١٠

ك<sub>ر</sub> : ١ ١ ١ ١ ١ ٢ ٢ ٥ ٦ ٤

( ٦ ) الأعداد الآتية هي الزيادة في الوزن بالكيلوجرامات لمجموعة من ١٣ بقرة بعد فترة من نظام غذائي معين :

٤,٢ ٤,٢ ٣,٢ ١,٢ ٢,٥ ١,٢ ٧,٢ ٥ ٩,٢ ٢,٢ ٨,٢ ٩,٢

أوجد الوسيط ونصف المدى الربيعي ومعامل الالتواء الربيعي .

## ملاحظة : استخدام الحاسبات :

تستطيع الحاسبات الالكترونية القيام بكل دقة وسرعة بالعمليات التي تتطلبها دراسة البيانات الاحصائية بدءا من تكوين الجداول والتوزيعات التكرارية من واقع البيانات الخام إلى حساب مقاييس النزعة المركزية والتشتت والالتواء والتفرطح .

## ( ٢ - ٥ ) شكل الساق والورقة : Stem - and - Leaf Diagram

إن أسلوب الأشكال المسماة بأشكال الساق والورقة هو تنوع لأسلوب التوزيعات التكرارية ، فهو يؤدي إلى تجميع أو تكثيف البيانات في عدد مناسب من الأقسام مثله في ذلك مثل التوزيعات التكرارية ، إلا أنه يتميز عنها بأمرين رئيسيين أولهما أنه يحتفظ بفردية كل عنصر من عناصر البيانات وثانيهما أنه يسهل لنا التعرف على البيانات وتكوين فكرة عن توزيع المتغير الذي تعبر عنه هذه البيانات . ولهذا يعتبر هذا الأسلوب من الأدوات الأولية المفيدة في عملية التحليل الاستطلاعي للبيانات .

وإنشاء شكل ساق وورقة هو أمر غاية في السهولة ، ولا يتطلب إلا فصل أرقام كل عدد في البيانات إلى جزئين أحدهما يسمى ساق والآخر يسمى ورقة . والمتباد أن تؤخذ الورقة على أنها الرقم الأخير في العدد أي الرقم الذي في أقصى يمينه ، أما الساق فهي بقية الأرقام . وتوضع الأرقام التي ترمز إلى السيقان رأسيا ثم توضع الأوراق المصاحبة لكل ساق أفقيا كما في الأمثلة الآتية .

## مثال ( ٢ - ٩ )

لأعداد الآتية هي : الأعمار عند حدوث صدمة قلبية لعينة من أربعين مريضا .  
أنشئ شكل ساق وورقة واذكر ملاحظاتك عنه .

٨١	٦٩	٤٢	٥٨	٧٨	٧٣	٣٧	٤٠	٥٨	٣١
٥٢	٦٠	٦١	٦٠	٢٩	٣١	٤٨	٤٩	٨١	٦٢
٥٢	٥٨	٥٢	٦٠	٦١	٧٣	٥٧	٣٤	٥٢	٦١
٧٧	٥٩	٤٥	٤٥	٤٤	٤٠	٥١	٣٧	٥٢	٤٤

الحل :

في هذا المثال الورقة هي الرقم الأخير في العدد وهو رقم الآحاد أما الساق فهي رقم العشرات . فمثلا للعدد ٣١ الواحد هو الورقة والثلاثة هي الساق ، وللعدد ٦٢ الاثنين هي الورقة والستة هي الساق وهكذا .

٢	٩	١
٣	١ ٤ ٧ ٧ ١	٥
٤	٤ ٠ ٩ ٨ ٠ ٤ ٥ ٢ ٥	٩
٥	٨ ٢ ٢ ٧ ١ ٨ ٢ ٨ ٩ ٢ ٢	١١
٦	٢ ١ ١ ٠ ٠ ١ ١	٨
٧	٣ ٣ ٨ ٧	٤
٨	١ ١	٢

الشكل (٢-٧)

شكل ساق وورقة للأعداد التي حدثت عندها صلعت قلية

لعينة من ٤٠ مريضاً

لإنشاء الشكل نبدأ بتحديد السيقان وهي أرقام العشرات فنجد أنها تتراوح بين ٨،٢ . نرسم خطاً رأسياً ثم نكتب الأرقام المثلثة للسيقان على يساره مرتبة ترتيباً تصاعدياً ونكون بذلك قد كتبنا جميع أرقام العشرات الممكنة ، ثم نمر على البيانات واحداً واحداً

لنكتب الرقم الممثل للورقة ( أي رقم الآحاد ) في كل منها في الصف الذي يناسبه أي على يمين الساق التي ترمز إلى رقم العشرات فيه . انظر الشكل ( ٢ - ٧ ) . نلاحظ أنه في الصف الأول من الشكل يوجد عدد واحد فقط وهو يمثل العمر ٢٩ ، وفي الصف الثاني توجد خمسة أعداد هي الأعمار ٣١ ، ٣٤ ، ٣٧ ، ٣٧ ، ٣١ وهكذا . وإذا جمعنا الأعداد التي بالصفوف جميعها لوجدناها مساوية للعدد ٤٠ وهو حجم العينة .

### ملاحظات :

١ - في بناء شكل الساق والورقة ينبغي أن نختار عددا مناسباً من السياق ، وذلك لكي نستطيع الإفادة من الشكل ، والمعاد ألا يقل هذا العدد عن خمسة وألا يزيد عن عشرين . هذا مع ملاحظة أن شكل الساق والورقة لا تكون له فائدة كبيرة إذا كان عدد البيانات كبيراً جداً أو كان المدى الذي يتغير فيه المتغير كبيراً .

٢ - يمكننا دائماً استعادة البيانات الأصلية من الشكل المرسوم وذلك بضم الساق مع كل ورقة من أوراقه ، وهذا مانعنا من القول أن الشكل يحفظ بفرديته البيانات .

٣ - لتسهيل التعرف على خصائص توزيع البيانات ، ندير الشكل بحيث يصبح الخط الرأسى أفقياً وتكون الأرقام الممثلة للسياق أسفل هذا الخط ، وبمساعداً في ذلك أيضاً أن نرسم خطاً ناعماً حول نهايات الأوراق ، ثم نحاول الإجابة عن تساؤلات كالآتية :

( أ ) هل تميل البيانات إلى التجمع حول ساق أو سياق معينة أم تتوزع على كل السياق بشكل متبادل ؟

( ب ) هل تشتت البيانات تشتتاً واسعاً أم ضيقاً ؟

( ج ) هل هناك تماثل في توزيع البيانات ؟ هل تميل البيانات إلى التناقص تدريجياً نحو أحد طرفي التوزيع ؟ هل هناك سمات خاصة يشير إليها المنحنى المرسوم حول نهايات الأوراق ؟

ففي المثال ( ٢ - ٩ ) نجد أن عددا كبيرا من البيانات ( ١١ عمرا ) يتجمع حول الساق ( ٥ ) ونلاحظ أن الشكل يكاد يكون متائلا حول هذه الساق ، كما نلاحظ أن الصدمات القلبية في هذه العينة يحدث أغلبها في الخمسينات ثم في الأربعينات والستينات ، وأن هذه الصدمات لا تحدث تقريبا قبل سن الثلاثين أو بعد سن الثمانين .

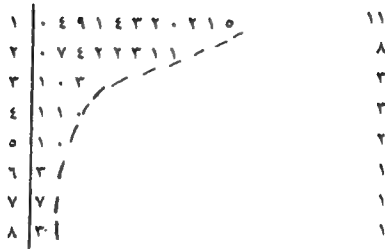
#### مثال ( ٢ - ١٠ )

ارسم شكل ساق وورقة للبيانات الآتية التي هي مشاهدات عن المتغير العشوائي الذي يعبر عن شدة الزلازل التي حدثت في أحد المناطق مقاسة بمقياس ريختر Richter. علق على الشكل .

١,٠	٨,٣	٣,١	١,١	٥,١	١,٢	١,٠	٤,١	١,١	٤,٠
٢,٠	١,٩	٦,٣	١,٤	١,٣	٣,٣	٢,٢	٢,٣	٢,١	٢,١
١,٤	٢,٧	٢,٤	٣,٠	٤,١	٥,٠	٢,٢	١,٢	٧,٧	١,٥

#### الحل :

في أي عدد في هذه البيانات الورقة هي الرقم الذي في خانة الجزء من عشرة ، والساق هي رقم الآحاد . تتراوح السيقان بين ١ ، ٨ وإذن نكتب الأرقام ١ ، ٢ ، ٨٠٠٠ على يسار نخط الرأس ثم نسجل الأوراق كل أمام الساق المناسبة فنحصل على الشكل ( ٢ - ٨ ) الآتي حيث نلاحظ أن هناك ١١ عددا في الصف الأول هي : ١,٠ ، ١,٤ ، ١,٩ ، ١,١ ، ١,٤ ، ١,٣ ، ١,٢ ، ١,٠ ، ١,٢ ، ١,١ ، ١,٥ ، ١,٠ وثمانية أعداد في الصف الثاني هي : ٢,٠ ، ٢,٧ ، ٢,٤ ، ٢,٢ ، ٢,٢ ، ٢,٣ ، ٢,١ ، ٢,١ وهكذا .



٣٠

#### الشكل (٢-٨)

شكل ساق وورقة لشدة الزلازل مقاسة بمقياس ريختر  
في عينة مأخوذة من أحد المناطق

#### التعليق :

تشتت شدة الزلازل بين القيمتين ١,٠ ، ٨,٣ غير أن البيانات تميل إلى التجمع حول القيم الصغرى وتقل تدريجياً في اتجاه القيم الكبرى ( هناك التواء إلى اليمين ) وهذا يعنى أن معظم الزلازل في هذه العينة كانت خفيفة . وإذا كانت هذه العينة تعبر عن المجتمع ككل ، فإن وقوع زلازل شديدة في هذه المنطقة يكون أمراً بعيد الاحتمال .

#### مثال ( ٢-١١ )

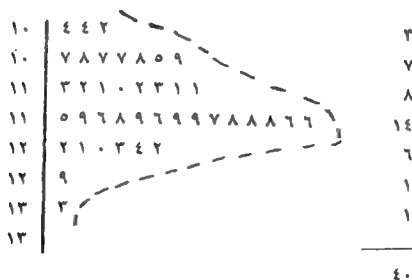
ارسم شكل ساق وورقة مع التعليق لبيانات المثال ( ٢-٢ ) السابق عن أطوال محيطات رؤوس عينة من ٤٠ من الحمام المنزلي وهي :

١١,٦ ١١,٨ ١١,٨ ١٢,٢ ١٢,٣ ١١,١ ١١,٩ ١١,٨ ١٢,٩ ١٢,٢  
١١,١ ١٠,٥ ١١,٢ ١٣,٣ ١١,٩ ١١,٦ ١١,٢ ١١,٣ ١١,٥ ١٠,٧

١١,٦ ١٠,٩ ١٠,٢ ١١,٩ ١١,٠ ١٠,٨ ١٠,٧ ١٠,٤ ١١,٩ ١٢,١  
١١,١ ١١,٣ ١١,٨ ١١,٧ ١٢,٤ ١٢,٠ ١٠,٧ ١٠,٤ ١١,٦ ١٠,٨

الحل :

في هذا المثال الورقة هي الرقم الذي بخانة الجزء من عشرة ، والساق هي العدد المكون من رقمي الآحاد والعشرات ، فمثلا للعدد ١٢,٢ الورقة هي ٢ والساق هي ١٢ . إذا استخدمنا الطريقة التي استخدمناها في المثالين السابقين نجد أن لدينا أربعة سيقان فقط هي ١٠ ، ١١ ، ١٢ ، ١٣ ويكون من الصعب تكوين فكرة عن التوزيع بهذا العدد القليل من السيقان . وللتغلب على هذه الصعوبة نأخذ كل رمز يرمز إلى ساق مرتين وبذلك نكون قد قسمنا البيانات إلى ٨ أقسام وهذا عدد مناسب ، على أن نكتب أمام الرمز في المرة الأولى الأوراق ٠ ، ١ ، ٢ ، ٣ ، ٤ ونكتب أمام الرمز في المرة الثانية الأوراق ٥ ، ٦ ، ٧ ، ٨ ، ٩ كما في الشكل ( ٢-٩ ) الآتي :



الشكل ( ٢-٩ )

شكل ساق وورقة لأطوال محطات رؤوس عينة من ٤٠ من الحمام الفزلي



## التعليق :

تشنت أطوال محيطات رؤوس الحمام بين القيمتين ١٠,٢ ، ١٣,٣ من المليمترات ، إلا أن عددا كبيرا منها يتجمع حول الساق ١١ ويقل هذا التجمع تدريجيا من الناحيتين بمقادير متوازنة تكاد تجعل الشكل متائلا .

## تمارين ( ٢ - ٢ )

١ - الآتي هي أعداد النقاط التي حاز عليها ٤٠ لاعبا في فرق كرة القدم بإحدى المدارس الثانوية . انشئ شكل ساق وورقة مستخدما السيقان ٠ ، ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ ، ٧ ثم علق على الشكل .

٠	٢	٥٩	٤٥	٠	١	١٨	٠	٥٧	٠
٢	٦٠	٧	٤	١	٠	٤٢	٩	٣	٣
٤٨	٣	٤٦	٨	٣	٦	٣٧	٣٥	١٧	٤
٧٥	٧٥	٢	٢١	٦	٢١	٢	٠	١٨	٢

٢ - أجريت دراسة لمدة تأثير التدخين على نمط النوم . المتغير العشوائي - الذي يدرس هو الزمن بالدقائق الذي يمضي حتى ينام الشخص ، وقد وجدت البيانات الآتية في عيتين عشوائيتين إحداهما من المدخنين والأخرى من غير المدخنين . المطلوب رسم شكل ساق وورقة لكل عينة باستخدام الأعداد من ١٥ إلى ٢٥ كسيقان ثم بيان ما إذا كان هناك فرق بين توزيع المتغير - في العيتين .

غير المدخنين								المدخنون							
١٧,٢	١٩,٧	١٨,١	١٥,١	١٨,٣	١٧,٦	١٥,١	٢٠,٥	١٧,٧	٢١,٣	١٦,٠	٢٤,٨	١٦,٢	١٩,٩	١٨,١	٢٢,١
١٦,٢	١٩,٩	١٩,٨	٢٣,٦	٢٤,٩	٢٠,١	١٦,٨	٢١,٢	١٨,١	٢٢,١	١٥,٩	٢٥,٢	١٦,٢	١٩,٩	١٨,١	٢٢,١
١٩,٨	٢٢,٦	٢٠,٠	٢٤,١	٢٥,٠	٢١,٤	٢٢,٨	٢٧,٤	١٩,٤	٢٥,٢	١٨,٣	٢٥,٠	١٦,٢	١٩,٩	١٨,١	٢٢,١
٢١,٢	١٨,٩	٢٧,١	٢٠,٦	٢٣,٣	٢٠,٢	٢٥,٨	٢٤,١	١٥,٠	٢٤,١	٢١,٦	١٦,٣	١٦,٢	١٩,٩	١٨,١	٢٢,١
٢١,١	١٦,٩	٢٣,٠	٢٠,١	١٧,٥	٢١,٣	٢٤,٣	٢٥,٧	١٥,٢	١٨,٠	٢٣,٨	١٧,٩	١٦,٢	١٩,٩	١٨,١	٢٢,١
٢١,٨	٢٢,١	٢١,١	٢٠,٥	٢٠,٤	٢٠,٧	٢٣,٢	٢٥,١	١٦,١	١٧,٢	٢٤,٩	١٩,٩	١٦,٢	١٩,٩	١٨,١	٢٢,١
١٩,٥	١٨,٨	١٩,٢	٢٢,٤	٩١,٣	١٧,٤	١٥,٧	١٥,٣	١٩,٩	٢٣,١	٢٣,٠	٢٥,١	١٦,٢	١٩,٩	١٨,١	٢٢,١

٣ - في تجربة نفسية عن التعلم استخدم ٤٠ فأراً قسموا عشوائياً إلى قسمين متساويين في العدد وأتيح لكل فأر أن يجري في متاهة وسجل الوقت الذي يستغرقه في إتمامها بالثواني . درست واحدة فقط من المجموعتين على الجري في المتاهة ، ثم أتيح لكل فأر أن يجري في المتاهة مرة ثانية وسجل الوقت الذي يستغرقه في هذه المرة الثانية . المتغير الذي يدرس هو الفرق في الوقت بين المراتين ( الوقت في المرة الأولى - الوقت في المرة الثانية ) . دونت هذه الفروق في الجدول الآتي :

الفئران غير المدربة				الفئران المدربة			
٢,١-	٢,٢-	١,١-	٢,٥-	٤,٠	٣,٢	٤,١	٤,٩
١,٢-	٢,٠	٢,٤-	٠,٦-	٤,٢	٣,٧	٤,٣	٤,٢
١,٣	١,٣-	٠,٢-	٢,٧-	٤,٤	٣,٦	٣,٥	٤,٩
١,٤	٠,٩	٢,٢	٢,١	٥,١	٤,٥	٤,٧	٥,٠
١,٨	٢,١	١,١	٢,٦	٥,٦	٤,٦	٥,٢	٥,٥

انشئ شكل ساق وورقة لكل من المجموعتين ( استخدم السيقان ٣ ، ٣ ، ٤ ، ٤ ، ٥ ، ٥ للفقران المدربة ؟ -٢ ، -١ ، -٠ ، ١٠٠ ، ٢ للفقران غير المدربة ) ثم حاول الاجابة عما يأتي :

( أ ) متى تنتج القيم الموجبة وماذا تعنى هذه القيم ؟

( ب ) من الواضح أن متوسط الفروق في عينة الفقران المدربة ( وجميعها موجبة ) أكبر منه في عينة الفقران غير المدربة ، فهل هذا يعنى أن الفقران تتعلم من التدريب ؟

( جـ ) قارن بين التوزيعين مستخدما شكلي الساق والورقة .



## الفصل الثالث

### بعض نماذج الاحتمال

#### SOME PROBABILITY MODELS

كما أشرنا في نهاية البند (١ - ٣) ، نصف المتغير  $x$  بأنه متغير عشوائي إذا كانت قيمه أعداداً حقيقية يتأثر قياسها بعوامل عشوائية ويكون ظهورها مصحوباً باحتمالات محددة . ولكل متغير عشوائي توزيع يربط بين قيمه واحتمالاتها يسمى بتوزيع الاحتمال لهذا المتغير . وتنقسم توزيعات الاحتمال للمتغيرات العشوائية إلى نوعين هما توزيعات الاحتمال الوثابة وتوزيعات الاحتمال المتصلة بحسب كون المتغير من النوع الوثاب أو من النوع المتصل .

#### (٣ - ١) توزيعات الاحتمال الوثابة

#### DISCRETE PROBABILITY DISTRIBUTIONS

ليكن  $x$  متغيراً حقيقياً وثاباً يأخذ القيم  
 $x_1, x_2, x_3, \dots$  باحتمالات قدرها  
 $p_1, p_2, p_3, \dots$  حيث  $\sum p_i = 1$   
إن تجمع قيم  $x$  مع احتمالاتها المناظرة في مجموعة من الأزواج المرتبة  
 $\{(x_1, p_1), (x_2, p_2), (x_3, p_3), \dots\}$   
يسمى بتوزيع الاحتمال للمتغير العشوائي  $x$  .

وفي كثير من الأحيان يمكن أن نجد دالة غير سالبة د حيث :

$$(1) \quad d(s) = l(s) = 0$$

أى حيث تكون قيمة الدالة د عند س مساوية لاحتمال أن يأخذ المتغير س القيمة س . وتسمى هذه الدالة حينئذ بدالة كتلة الاحتمال للمتغير س probability mass function

مثال (٣ - ١) :

التوزيع الآتي هو توزيع الاحتمال لعدد مرات إصابة الهدف لجندى ما يطلق بندقته على هدف ثابت ٥ مرات .

عدد مرات إصابة الهدف س : صفر ١ ٢ ٣ ٤ ٥  
احتمال هذا العدد ل : صفر ٠,٤ ٠,٣ ٠,٢ ٠,١ ٠,١ صفر  
لاحظ أن  $0.1 + 0.1 + 0.2 + 0.3 + 0.4 + 0.1 = 1$

ويمكن هنا التعبير عن الاحتمالات بواسطة الدالة د المعرفة كالآتي :

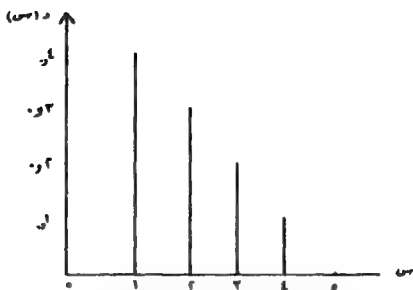
د (س) =  $\begin{cases} 0.1 & \text{حيث س = ١, ٢, ٣, ٤, ٥} \\ 0 & \text{فيما عدا ذلك} \end{cases}$  صفر

إذ أنه بوضع س = ١, ٢, ٣, ٤, ٥ نحصل على الاحتمالات ٠,١, ٠,٢, ٠,٣, ٠,٤, ٠,١ صفر .

ونمثل هذه الدالة أو هذا التوزيع بيانيا كما في الشكل (٣ - ١) الآتي .

### (٣-١-١) الوسط الحسابي والتباين : MEAN AND VARIANCE

إن توزيعات الاحتمال الوثابة تشبه التوزيعات التكرارية ، إلا أن الاحتمالات ل تحمل محل التكرارات كـ ، كما أن حجم التوزيع هو الواحد الصحيح دائماً . وهذا الواحد يشير إلى أن هناك احتمالاً قدره الواحد الصحيح موزعاً على القيم المختلفة للمتغير ولذلك سمى التوزيع بتوزيع الاحتمال .



الشكل (٣-١) الأعمدة البائية لتوزيع احتمالات  
عدد مرات إصابة الهدف لجندى يطلق على الهدف خمس طلقات

ومن هنا كان تعريفا الوسط الحسابي والتباين لتوزيع احتمال - وسنرمز لهما بالرمزين  $\mu$  ،  $\sigma^2$  ، يشبهان تعريفى الوسط الحسابي والتباين للتوزيع التكرارى فهما يعرفان كالآتي :

(٢) الوسط الحسابي  $\mu =$  مح لـ سـ

التباين  $\sigma^2 =$  مح لـ  $(س - \mu)$

(٣) أو  $\sigma^2 =$  مح لـ  $س^2 - \mu^2$

أما الانحراف المعياري فهو بالطبع الجذر التربيعي للتباين .

ففي المثال (٣-١) السابق نجد أن :

$$\mu = \text{مح لـ س} =$$

$$= 0 \times 0 + 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.3 + 4 \times 0.4 + 0 \times 0 =$$

$$\sigma^2 = \text{مح لـ س}^2 - \mu^2 =$$

$$= (0 \times 0 + 1 \times 0.1 + 4 \times 0.2 + 9 \times 0.3 + 16 \times 0.4 + 0 \times 0) -$$

$$= 4 - 0 = 4$$

## PROBABILITY MODELS

### (٣-٢) نماذج الاحتمال :

نموذج الاحتمال للمتغير عشوائي  $x$  هو توزيع احتمال ذو صيغة رياضية محددة يفترض أنها تعكس سلوك المتغير  $x$ . ويميز عن الاحتمالات من هذا النموذج بدلالة واحد أو أكثر من أدلة مجهولة تتوقف على خواص المجتمع وطريقة المعاينة منه. ويبنى كل نموذج احتمال على افتراضات خاصة تصور تصويراً مناسباً الميكانيكية العشوائية التي تسبب الاختلافات في مشاهدتنا عن المتغير.

وستتناول فيما يلي أربعة من أشهر نماذج الاحتمال للمتغيرات العشوائية الوثابة تعرف بتوزيع ذي الحدين وتوزيع بواسون وتوزيع باسكال والتوزيع الهندسي.

### (٣-٣) توزيع ذي الحدين : THE BINOMIAL DISTRIBUTION

في كثير من الأحيان يكون اهتمامنا منصباً على وجود أو عدم وجود خاصية ما في وحدات أو عناصر مجتمع ما. ولذلك ننظر إلى المجتمع على أنه مقسم إلى قسمين منفصلين بحسب هذه الخاصية. فمثلاً قد نقسم مجتمعاً من الطلاب بحسب خاصية القومية إلى القسمين : عربي / غير عربي أو بحسب الجنس إلى : ذكر / أنثى أو إلى القسمين : يدخن / لا يدخن أو إلى القسمين : يلبس نظارة / لا يلبس نظارة أو إلى أى قسمين منفصلين متكاملين.

في مثل هذه الحال تتركز الصفة الأساسية للمجتمع في دليل أو بارامتر واحد هو نسبة أى من القسمين - وليكن القسم الأول - إلى المجتمع كله ، وسنرمز إلى هذه النسبة بالرمز  $H$ . فمثلاً قد تكون  $H$  نسبة الطلاب العرب في المجتمع ، وهنا تكون نسبة الطلاب غير العرب هي  $1 - H$  مثلاً. وهذا يعني أننا إذا سحبنا عشوائياً عنصراً من المجتمع فإن  $H$  تعبر عن احتمال أن يكون هذا العنصر من القسم الأول ( طالب عربي مثلاً ) كما أن  $1 - H$  تعبر عن احتمال أن يكون العنصر من القسم الثاني ( طالب غير عربي ) .



سنسمى ظهور عنصر من القسم الأول نجاحاً للمحاكمة أو الحدث الذي ندرسه  
وظهور عنصر من القسم الثاني فشلاً للحدث ، أى سنعتبر أن لدينا حدثاً واحداً -  
مثلا ظهور طالب عربي - إما أن يقع أو لا يقع .

إن هدفنا الأساسي من هذه الدراسة يتلخص فيما يلي : نفرض أننا مسحنا من  
المجتمع عينة عشوائية حجمها  $n = 10$  . مثلاً . هناك ١١ حالة ، إذ يمكن أن تكون  
هذه العينة خالية من أى عنصر من القسم الأول كما يمكن أن تشتمل على عنصر  
واحد فقط من هذا القسم أو تشتمل على عنصرين أو ثلاثة أو أربعة أو ... عشرة .  
أى أن عدد مرات نجاح الحدث يمكن أن يكون ٠ ، ١ ، ٢ ، ٣ ، ... ، ١٠ .  
والسؤال الذى نستهدف الإجابة عنه هو : ما احتمال كل من هذه الحالات ؟ وبمعنى  
آخر إذا اعتبرنا أن لدينا متغيراً عشوائياً  $X$  يعبر عن عدد مرات نجاح الحدث  
فما توزيع الاحتمال لهذا المتغير ؟

إن الإجابة عن هذا السؤال تتوقف على ما نضعه من افتراضات نرى أنها مناسبة  
لما يهمنا من أوضاع ويمكن تحققها في عملية التجريب . وسوف نتبنى هنا  
الافتراضات أو الشروط الآتية :

#### (١) عشوائية العينة :

سنفترض أن المعاينة ( أى سحب العناصر من المجتمع ) عشوائية .

#### (٢) ثبات الدليل ح :

سنفترض أن احتمال نجاح الحدث هو عدد ثابت ح طوال عملية سحب العينة .  
ولكى يتحقق هذا الفرض سنعتبر أن حجم المجتمع كبيراً بالنسبة لحجم العينة وبالتالي  
فإنه حتى إذا كانت المعاينة بغير إرجاع يكون التغير الذى يحدث في قيمة ح تغيراً  
طفيفاً يمكن التجاوز عنه ، كما سنفترض أيضاً أن قيمة ح لا تختلف من عينة إلى  
أخرى .

### (٣) استقلال الأحداث :

سنفترض أن نجاح ( أو فشل ) الحدث في أى سحبة مستقل عما نتج من نجاح أو فشل في السحبات السابقة ، أى أن ما تسفر عنه أى سحبة لا يتأثر بأى حال بما نتج في السحبات الأخرى . كما سنفترض أن عدد مرات نجاح الحدث في عينة ما مستقل عن عدد مرات نجاحه في أى عينة أخرى .

نحت هذه الشروط وباستخدام قاعدة الضرب للأحداث المستقلة وقاعدة الجمع للأحداث المتنافية - انظر البند (١ - ٧) عن توافقات الاحتمال - نستطيع أن نثبت رياضياً أنه إذا كان المتغير  $x$  يعبر عن عدد مرات نجاح الحدث فإن دالة كتلة احتماله تأخذ الصورة الآتية :

$$d(x) = \binom{n}{x} p^x q^{n-x} \quad \text{حيث } x = 0, 1, \dots, n \quad (4)$$

وحيث  $0 < p < 1$  ،  $q = 1 - p$  ،  $n$  هو حجم العينة .

ويسمى توزيع الاحتمال حيثخذ بتوزيع ذى الحدين .

كما نستطيع أن نثبت رياضياً أن

$$(5) \quad \text{الوسط الحسابي للتوزيع} = np$$

$$(6) \quad \text{تباين التوزيع} = npq$$

وجدير بالملاحظة أن الدالة (٤) تتوقف على اثنين من الأدلة هما  $n$  ،  $p$  ومعرفة هذين الدليلين تحدد التوزيع تحديداً تاماً . ولذلك سنرمز لتوزيع ذى الحدين بالرمز  $(n, p)$  .

ملاحظة :

يمكن إيجاد التوافقات  $\binom{n}{x}$  بالطريقة الحسابية المعتادة أو باستخدام مثلث

باسكال أو باستخدام جداول جاهزة حسبت فيها هذه التوافقات - انظر الجدول

(٢) في ملحق هذا الكتاب وهو يعطى التوافقات  $\bar{u}$  لبعض قيم  $n$ ،  $s$  (  $n = ١، ٢، ٣، \dots، ٢٠$  و  $s = ٠، ١، ٢، \dots، ٢٠$  ) كما يمكن إيجاد قيم الاحتمالات (٤) من الجدول (٣) لبعض قيم  $n$ ،  $h$  (  $n = ٢، ٣، \dots، ١٥$  و  $h = ٠، ١، ٢، ٣، ٤، ٥، ٦، ٧، ٨، ٩، ١٠، ١١، ١٢، ١٣، ١٤، ١٥$  ) .

مثال (٣ - ٢) :

افرض أن احتمال ولادة مولود ذكر هو  $٠,٥$  واعتبر العائلات التي أنجبت ٤ أطفال .  
( أ ) أوجد توزيع احتمال المتغير  $x$  الذى يعبر عن عدد الذكور في العائلات ذوات الأربعة الأطفال .

(ب) إذا أخذنا عشوائياً ٢٠٠٠ عائلة من هذا النوع فما العدد الذى نتوقعه للعائلات التي يكون بها ولدان على الأقل ؟

الحل :

لدينا مجتمع من الولادات مقسم إلى القسمين ذكر / أنثى ، واحتمال وقوع أو نجاح الحدث « المولود ذكر » هو عدد ثابت  $h = \frac{1}{2}$  . إن المتغير  $x$  يعبر هنا عن عدد الأولاد ( الذكور ) في العائلات ذوات الأربعة الأطفال أى عدد المرات التي تحدث فيها ولادة مولود ذكر في هذا النوع من العائلات وإذن فالمتغير  $x$  لا يأخذ إلا القيم  $٠، ١، ٢، ٣، ٤$  . إن كل عائلة تعتبر عينة عشوائية حجمها  $n = ٤$  فإذا فرضنا أن إنجاب مولود ذكر ( أو أنثى ) في أى ولادة مستقل عما ينتج في الولادات الأخرى تكون الافتراضات الثلاثة لتوزيع ذى الحدين متوفرة ويكون للمتغير  $x$  توزيع ذى الحدين دليلاً  $n = ٤$  ،  $h = \frac{1}{2}$  وبالتالي تكون دالة كتلة احتماله :

$$D(s) = \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right) = \frac{1}{16}$$

حيث  $s = 0, 1, 2, 3, 4$

(أ) توزيع الاحتمال المطلوب هو التوزيع المين في العمودين الأول والثاني من الجدول (٣ - ١) الآتي :

الجدول (٣ - ١)

توزيع الاحمال لعدد الذكور في العائلات ذوات الأربعة الأطفال -  $h = 0, 1, 2, 3, 4$

عدد الذكور في العائلة	احتمال هذا العدد	العدد المتوقع من العائلات
$s$	$D(s)$	$2 \times D(s)$
0	$\frac{1}{16} = \frac{1}{2^4}$	125
1	$\frac{3}{16} = \frac{3}{2^4}$	500
2	$\frac{6}{16} = \frac{3}{8}$	750
3	$\frac{3}{16} = \frac{3}{2^4}$	500
4	$\frac{1}{16} = \frac{1}{2^4}$	125
	1,000	2,000

(ب) أما العمود الثالث فيعطى الأعداد المتوقعة من العائلات التي بها 0 ، 1 ، 2 ، 3 ، 4 أولاد من بين الـ 2,000 عائلة ، ومن هذا العمود ينتج أن العدد المتوقع للعائلات التي بها ولدين على الأقل .

= العدد المتوقع للعائلات التي بها ولدين أو ثلاثة أو أربعة .

$$= 125 + 500 + 750 = 1375 \text{ عائلة .}$$

مثال (٣ - ٣):

حسب نظرية مندل للخواص الوراثية ، حين يهجن نوع من النباتات حمراء

الزهور مع نوع ذى قرابة من نباتات يبيضاء الزهور تنتج خلفه ٢٥٪ منها حمراء الزهور . نفرض أننا سنقوم بتهجين ٥ أزواج من هذه النباتات نختارها عشوائياً فما احتمال أن يكون من بين خمسة السلالات الناتجة :

( أ ) لا توجد نباتات حمراء الزهور .

(ب) يوجد على الأقل ٤ نباتات حمراء الزهور ؟

الحل :

لدينا مجتمع من السلالات مقسم إلى القسمين : حمراء الزهور / يبيضاء الزهور . واحتمال الحدث « حمراء الزهور » هو عدد ثابت  $H = 0,25$  ومن الطبيعي أن نعتبر أن النواتج مستقلة لأن التهجين يتم بين أزواج مختلفة من النباتات . وعلى ذلك تكون الشروط الثلاثة متوفرة ويكون لدينا متغير عشوائي  $X$  يعبر عن عدد النباتات حمراء الزهور في العينات العشوائية ذوات الحجم  $n = 5$  وهو متغير له توزيع ذى الحدين دليلاً  $0, 25, 5$  ودالة كتلة احتمال

$$D(s) = \binom{n}{s} p^s (1-p)^{n-s} \quad \text{حيث } s = 0, 1, 2, 3, 4, 5$$

( أ ) احتمال عدم وجود نباتات حمراء الزهور .

$$L \quad (s = 0) = D(0) = \binom{n}{0} p^0 (1-p)^{n-0} = \binom{5}{0} (0,25)^0 (0,75)^5 = 0,234$$

$$= 0,234$$

(ب) احتمال وجود ٤ نباتات حمراء الزهور على الأقل .

$$L = (s \leq 4) = D(4) + D(5)$$

حدثان متنافيان

$$= \binom{5}{4} (0,25)^4 (0,75)^1 + \binom{5}{5} (0,25)^5 (0,75)^0$$

$$= 0,016$$

### (٣-٣-١) تقدير الدليل ح :

في توزيع ذى الحدين ، إذا كانت قيمة البارامتر ح مجهولة ، يمكن تقديرها تجريبياً من عينات بحيث تتوفر الشروط الثلاثة سالفة الذكر . فإذا حصلنا على التوزيع التكرارى لعينة حجمها ن ووجدنا أن وسطها الحسابي  $\bar{x}$  فإننا نأخذ هذا الوسط كتقدير للوسط الحسابي لتوزيع ذى الحدين وهو كما نعلم يساوى ن ح وبالتالي نقدر البارامتر ح بالعدد وحيث :

$$\bar{x} = r \quad (٧)$$

### مثال (٣-٤) :

لتقدير نسبة الحصى الجرانيتية إلى مجتمع الحصى على أحد الشواطئ أخذت من هذا الشاطئ ١٠٠ عينة عشوائية بكل منها ثلاث حصوات وحسب عدد الحصوات الجرانيتية بكل عينة فنتج التوزيع التكرارى الآتى :

عدد الحصوات الجرانيتية في العينة	٠	١	٢	٣
عدد العينات	٥٨	٣٣	٧	٢
الوسط الحسابي لهذا التوزيع التكرارى	$\bar{x} = \frac{1}{n} \sum x_i \cdot K_i$			
	$= \frac{1}{100} (0 + 33 + 14 + 6) = 0,53$			

إذن تقدير البارامتر ح من العينة هو  $r = \bar{x} = \frac{0,53}{3} = 0,177$  تقريباً

هذا على فرض أن توزيع عدد الحصى الجرانيتية هو توزيع ذى الحدين دليله ن ، ح حيث ن = ٣ ، ولنا أن نقول حيث أن هناك حوالى ١٧,٧٪ من الحصى الجرانيتية في مجتمع الحصى الذى على ذلك الشاطئ . ويمكننا اختبار مدى صواب هذا القول كما في البند التالى .

ويلاحظ أنه يمكن أيضاً تقدير النسبة ح من عينة عشوائية واحدة بشرط أن تكون كبيرة الحجم وذلك بأخذ التكرار النسبي لعدد المحصى الجرائمية التي ظهرت في العينة . ففى هذا المثال لدينا ٣٠٠ حصوة منها ٥٣ حصوة جرائمية ( ٠ + ٣٣ + ١٤ + ٦ ) وعلى ذلك فالتكرار النسبي للمحصى الجرائمية هو  $\frac{٥٣}{٣٠٠} = ٠,١٧٧ =$

(٣ - ٣ - ٢) اختيار ما إذا كان الدليل ح له قيمة معينة . توفيق توزيع ذى الحدين لتوزيع تكرارى معلوم .

نفرض أن قائلاً ذكر أن الدليل ح لتوزيع ذى الحدين المتغير ما له قيمة معينة أمثلاً ونريد اختبار هذا القول . لتحقيق هذا الفرض تتبع الخطوات الثلاث الآتية :

(أ) نختار عينات عشوائية من حجم معين ن ونحسب عدد مرات وقوع الحدث في كل منها أى نحسب العدد س ( حيث س = ٠ ، ١ ، ٢ ، ... ، ن ) في كل منها . نجمع هذه البيانات في توزيع تكرارى نحصل على التكرارات ك<sub>١</sub> ، ك<sub>٢</sub> ، ... ، ك<sub>ن</sub> المناظرة للأعداد ٠ ، ١ ، ٢ ، ... ، ن . إن هذه التكرارات نرمز لها بالرمز ك<sub>١</sub> ونسميها بالتكرارات المشاهدة . إن التوزيع الناتج يكون على نمط التوزيع الوارد بالمثال (٣ - ٤) .

(ب) إذا كان القول أو الفرض ح = أ صحيحاً فإن توزيع الاحتمال للمتغير ذى الحدين يكون دليلاً ن ، أ معروفين ونستطيع إيجاد هذا التوزيع والحصول على الاحتمالات ل<sub>١</sub> ، ل<sub>٢</sub> ، ... ، ل<sub>ن</sub> المناظرة للأعداد ٠ ، ١ ، ٢ ، ... ، ن . نضرب كلا من هذه الاحتمالات ( التكرارات النسبية النظرية أو المتوقعة ) في عدد العينات المأخوذة نحصل على الأعداد ق<sub>١</sub> ، ق<sub>٢</sub> ، ... ، ق<sub>ن</sub> . إن هذه الأعداد نرمز لها بالرمز ق<sub>١</sub> ونسميها بالتكرارات النظرية أو المتوقعة .

(ج) نقارن بين التكرارات النظرية ق<sub>١</sub> والتكرارات المشاهدة ك<sub>١</sub> المناظرة لها فإذا كانت المطابقة حسنة أى كانت أزواج التكرارات قريبة من بعضها بدرجة

معقولة بحيث لا يكون بينها فروق كبيرة جاز لنا قبول الفرض أن  $H = A$  ولا نرفضه .

إن مثل هذا الاختبار يدخل في موضوع اختبارات الفروض الذى سنتناوله في فصل لاحق حيث ستتعرف على أدوات تمكنتنا من الحكم حكماً موضوعياً على مدى صغر أو كبير مثل هذه الفروق وبالتالي من الحكم على صواب أو خطأ ذلك الفرض .

مثال (٣ - ٥) :

لاختبار الفرض القائل أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال ، اختيرت عشوائياً ٣٢٠ عائلة بكل منها ٤ أطفال فتتج التوزيع التكرارى الآتى :

عدد الأطفال الذكور في العائلة سـ	٠	١	٢	٣	٤
عدد العائلات كـ	١٣	٤٣	٩٢	١١٢	٦٠

هل هذه البيانات تدعم الفرض المذكور أو تنفيه ؟

الحل :

على فرض أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال ، فإن احتمال ولادة مولود ذكر يكون  $H = \frac{1}{2}$  وإذا كان هذا الفرض صحيحاً يكون توزيع عدد الذكور في العائلات ذوات الأربعة الأطفال هو توزيع ذى الحدين دليلاً ٤ ،  $\frac{1}{2}$  . وكما في المثال (٣ - ٢) يتولد لدينا التوزيع الآتى :

عدد الأطفال الذكور في العائلة سـ	٠	١	٢	٣	٤
احتمال هذا العدد كـ	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{16}$

ونظراً لأن عدد العائلات في العينات المأخوذة ٣٢٠ فإن التكرارات النظرية التى يمكن مقارنتها بالتكرارات المشاهدة تتج بضرب هذه الاحتمالات في ٣٢٠ ونحصل بذلك على التوزيع التكرارى النظرى الآتى :



عدد الأطفال في العائلة من : ٠ ١ ٢ ٣ ٤  
العدد المتوقع للعائلات : ٢٠ ٨٠ ١٢٠ ٨٠ ٢٠ (المجموع ٣٢٠)

ونقول هنا أننا قد وقفنا توزيع ذى الحدين للتوزيع التكرارى المعطى .

بمقارنة التكرارات النظرية مع التكرارات المشاهدة كـ وهى :

كـ : ٢٠ ٨٠ ١٢٠ ٨٠ ٢٠  
كـ : ١٣ ٤٣ ٩٢ ١١٢ ٦٠ (المجموع ٣٢٠)

نجد أن هناك تفاوتاً كبيراً بينها . وعلى فرض توفر شرط العشوائية والاستقلال فإن هذا التفاوت يعزى إلى خطأ الفرض أن  $\chi^2 = \frac{1}{4}$  .

وينبغى أن نشير هنا مرة أخرى إلى أن حكمنا هذا هو حكم ذاتي قد لا يكون هو الحكم الصحيح ، أما الحكم الموضوعى فيستلزم اللجوء إلى أحد الاختبارات الإحصائية المناسبة مثل اختبار  $\chi^2$  الذى سندرسه بعد . انظر المثال (٦ - ٨) في البند (٦ - ٧ - ٢) .

### تمارين (٣ - ١)

١ - في نوع من أبصال الزهور المعروف أن معدل الإنبات ٩٥٪ . تبعاً هذه الأبصال وتباع في عبوات يحتوى كل منها على ١٠ أبصال . إذا سحب أحد هذه العبوات عشوائياً وزرع ما بها من أبصال فما احتمال كل من الحدثين الآتين :  
(أ) لا تنبت أى بصلة .

(ب) تنبت بصلة واحدة على الأقل ؟

٢ - معدل الإصابة بمرض ما في نوع من البقر ٢٥٪ اختبرت عينة عشوائية من ٨ بقرات . أوجد :

- (أ) احتمال أن تكون بقرتان بالضبط مصابتين .  
 (ب) الوسط الحسابي والتباين لعدد البقرات المصابة في العينات من الحجم ٨ .  
 لماذا ينبغي أن نفترض هنا أن هذا المرض غير معد للبقر ؟  
 ٣ - احتمال إصابة هدف ثابت ٠,٢ . إذا أطلقت ٥ طلقات مستقلة على هذا الهدف فما احتمال إصابته مرة واحدة على الأقل ؟

٤ - لاختبار الفرض القائل أن نسبة الحصى الجرانيتية إلى مجتمع الحصى على أحد الشواطئ هو  $0,02 =$  أخذت من هذا الشاطئ ١٠٠ عينة عشوائية بكل منها ثلاث حصوات وحسب عدد الحصوات الجرانيتية بكل عينة فنتج التوزيع التكراري المدون في المثال (٣ - ٤) وهو :

عدد الحصوات الجرانيتية في العينة :	٠	١	٢	٣
عدد العينات	٥٨	٣٣	٧	٢

اختبر ما إذا كانت هذه البيانات تدعم الفرض المذكور على فرض أن توزيع عدد الحصى الجرانيتية هو توزيع ذي الحدين .

### (٣ - ٤) توزيع بواسون : POISSON DISTRIBUTION

توزيع بواسون هو توزيع احتمال لتغير عشوائي وثاب سم تأخذ دالة كتلة احتمال الصورة

د (س) =  $\frac{e^{-\lambda} \lambda^s}{s!}$  حيث  $s = 0, 1, 2, \dots, \infty$  (أ)  
 وحيث  $\lambda$  أساس اللوغاريتمات الطبيعية ( $\lambda = 2,71828$  تقريباً)  
 وهذه الدالة دليل واحد هو العدد  $m$  وبالتالي يتحدد التوزيع تماماً إذا عرفت قيمة  $m$  . فمثلاً حين  $m = 2$  يكون التوزيع كالآتي :

س :	٠	١	٢	٣	٤	٥	٦	...
د (س) :	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2} e^{-\lambda}$	$\frac{\lambda^3}{6} e^{-\lambda}$	$\frac{\lambda^4}{24} e^{-\lambda}$	$\frac{\lambda^5}{120} e^{-\lambda}$	$\frac{\lambda^6}{720} e^{-\lambda}$	...
أى د (س) :	٠,١٣٥	٠,٢٧١	٠,٢٧١	٠,١٨٠	٠,٠٩٠	٠,٠٣٦	٠,٠١٢	...

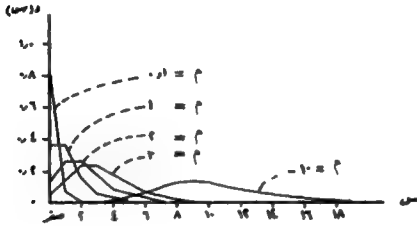
ومن المميزات الرئيسية لتوزيع بواسون أن

$$\text{الوسط الحسابي} = \text{التباين} = \mu$$

(٩)

ملاحظة (١) :

مجموعة قيم المتغير البواسوني  $\mu$  هي مجموعة لا نهائية  $\{0, 1, 2, \dots\}$  إلا أنه بعد قيمة معينة من تتوقف على الدليل  $\mu$  ، تتناقص احتمالات هذه القيم تدريجياً حتي تكاد تنعدم كما يشير إلى ذلك الشكل (٣ - ٢) الآتي :



الشكل (٣ - ٢) الصلحات التكرارية لتوزيع بواسون لقيم مختلفة للوسط الحسابي  $\mu$

ملاحظة (٢) :

هناك جداول تعطى قيم  $e^{-\mu}$  لبعض قيم  $\mu$  - انظر الجدول (٤) بملحق الكتاب - كما أن هناك جداول تعطى الاحتمالات والاحتمالات المتجمعة ل  $(\mu = s)$  ، ل  $(s \geq \mu)$  لتوزيع بواسون لبعض قيم الدليل  $\mu$  والعدد  $s$  - انظر الجدول (٥) بملحق الكتاب .

يستفاد من توزيع بواسون في الموضوعين المتقدمين بالبندين الآتين .

### (٣ - ٤ - ١) توزيع بواسون كتقريب لتوزيع ذى الحدين :

نستطيع رياضياً إثبات أنه بوضع  $m = n$  ح في توزيع ذى الحدين المعرف بالدالة (٤) فإنه يؤول إلى توزيع بواسون المعرف بالدالة (٨) حين تقترب  $n$  من اللانهاية وتقترب  $n$  من الصفر .

وهذا يعني من الناحية العملية أنه حين يكون حجم العينة  $n$  كبيراً والاحتمال الثابت  $n$  صغيراً فإن الاحتمالات في توزيع ذى الحدين يمكن إيجادها بالتقريب من الدالة (٨) بدلا من الدالة (٤) مع وضع  $m = n$  ح أى بحيث يكون متوسط توزيع بواسون مساوياً لمتوسط توزيع ذى الحدين . وقد وجد أن هذا التقريب يكون جيداً ، أى يمكن التجاوز عن الخطأ الناشئ عنه إذا كانت :

$$(n \leq 50, n \geq 5) \text{ أو } (n \geq 0.1, n \geq 5) \quad (10)$$

إن هذا التقريب من شأنه تبسيط حساب احتمالات ذى الحدين إذ أن حسابها من الدالة (٨) أسهل بكثير من حسابها من الدالة (٤) خاصة إذا كانت  $n$  خارج الحدود التي وضعت لها جداول ذى الحدين .

### مثال (٣ - ٦) :

إذا كان احتمال أن يحدث رد فعل سيء للشخص الذى يحقن بمصل ما هو ٠,٠٠١ فاحسب احتمال أن تحدث ٤ حالات ردود فعل سيئة من بين ٢٠٠٠ شخص يحقنون بهذا المصل .

### الحل :

توزيع عدد الأشخاص الذين يحدث لهم ردود فعل سيئة هو توزيع ذى الحدين دليلاً  $n = 2000$  ،  $p = 0.001$  ودالة كتلة احتماله هي :

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad \text{حيث } x = 0, 1, 2, \dots, 2000$$

احتمال ٤ حالات و حدود فعل سيئة هو :

$$ل (٣ = ٤) = د (٤) = ٠,٠٠١ = ٠,٠٠١ (٠,٩٩٩) = ٠,١٠٥٠$$

من الواضح أن حساب هذا الاحتمال لم يكن سهلاً فقد تطلب استخدام اللوغاريتمات والحاسب ، غير أنه يمكننا إيجاد الاحتمال المطلوب تقريباً من توزيع بواسون نظراً لتوفر أحد شروط التقريب وهو أن  $ن = ٢٠٠٠$  أكبر من ٥٠ ،  $ن ح = ٢$  أصغر من خمسة .

$$نضع م = ن ح = ٢٠٠٠ \times ٠,٠٠١ = ٢$$

حيث  $م = ٠, ١, ٢, \dots$

$$د (م) = \frac{e^{-م} م^م}{م!}$$

$$٠,٠٩٠٢ = ٠,١٣٥٣ \times \frac{٢}{٢!} = \frac{٢}{٢!} = ٠,١٣٥٣$$

وهذا الناتج يمكن إيجاده من الجدول (٥) وهو قريب جداً من الناتج السابق وهو ٠,١٠٥٠

### (٣ - ٤ - ٢) توزيع بواسون كنموذج لتوزيع الأحداث النادرة :

بصرف النظر عن الدور الذي يلعبه توزيع بواسون كتقريب لتوزيع ذي الحدين فإن له شخصيته واستخداماته الخاصة ، فقد دلت التجربة على أنه يصلح كنموذج احتمال لبعض الأحداث التي تقع عشوائياً عبر الزمان أو المكان .

وعلى سبيل المثال وجد أنه باختيار قيمة مناسبة للدليل م فإن توزيع المتغير م الذي يعبر عن عدد جسيمات ألفا التي تنبعث من مادة مشعة في وحدة زمن مناسبة يمكن أن يعتبر توزيعاً بواسونياً . وبالمثل للمتغيرات التي تعبر ( في فترات زمن مناسبة ) عن عدد التفجرات الفجائية في الصفات الوراثية - عدد حوادث الطائرات - تنابع طلبات الإغاثة على مراكز الإسعاف - تنابع المكالمات التليفونية على مراكز الهاتف - عدد ولادات ٤ توأم في مدينة - عدد حالات الانفلوانزا

التي ترد إلى مستشفى كبير ... كذلك المتغيرات التي تعبر عن عدد الطحالب في مربع على سطح جبل - عدد الطفيليات على أحد الموائل - عدد البكتريا من نوع معين على طبق بترى Petri plate - عدد الأخطاء المطبعية في صفحة كتاب - الحلال الذى يحدث في جهاز معقد ...

إن مثل هذه الحالات ينظر إليها غمطياً على أنها عملية تولد عدداً من التغيرات أو الأحداث ( مثل ظهور جسيم ألفا ، طحلب ، بكتريا ) في وحدة زمن أو وحدة فراغ مناسبة ، وهذه الوحدة مقسمة إلى عدد كبير جداً من الأجزاء الصغيرة جداً سنسميها لحظات instants ( سواء كان التقسيم من حيث الزمن أو من حيث الفراغ ) وكل من هذه اللحظات يعتبر محاولة يمكن أن يقع فيها الحدث أو لا يقع ، وبالتالي فإن هناك إمكانية وقوع الحدث في عدد كبير من المرات .

ويتخذ المتغير توزيعاً بواسونياً إذا توفر الشرطان الآتيان :

### ( أولاً ) ندوة الحدث :

يشترط أن يكون معدل وقوع الحدث ، أى متوسط عدد مرات وقوعه في وحدة الزمن أو وحدة الفراغ ، صغيراً بالنسبة لعدد المحاولات التي يمكن أن تسفر عن وقوع الحدث . وهذا ما يجعلنا نصف الحدث بأنه حدث نادر . اعتبر مثلاً توزيع المتغير  $x$  الذى يعبر عن عدد البكتريا في طبق بترى : إن وحدة الفراغ هنا هي طبق بترى الذى ننظر إليه على أنه مكون من عدد كبير من المساحات الميكروسكوبية ( لحظات ) كل منها قد يشتمل أو لا يشتمل على بكتريا . وجد بالتجربة أن متوسط عدد البكتريا في الطبق هو عدد متواضع بالرغم من أن هناك بالفعل عدداً لا نهائياً من المحاولات التي يمكن أن تنتج عنها بكتريا ، وعلى هذا فالحدث هو حدث نادر . كذلك اعتبر المتغير الذى يعبر عن عدد الأخطاء في صفحة كتاب جيد الطبع . إن هذه الصفحة تتألف من عدد كبير من الكلمات ( لحظات ) كل منها قد يقع في كتابتها خطأ أو لا يقع وعلى ذلك فهناك إمكانية

وقوع عدد كبير من الأخطاء ، ولكن نظراً لأن معدل وقوع الخطأ هو عدد صغير جداً نعتبر أن هذا الحدث هو حدث نادر .

واعتبار الحدث نادر هو مسألة نسبية تتطلب أن تكون وحدة الزمن أو وحدة الفراغ كبيرة كبيراً كافياً ، فمثلاً عندما نعد الطحالب على مربع ما يجب أن يكون هذا المربع كبيراً كبيراً كافياً يسمح بنمو عدد وافر من الطحالب مادامت الظروف البيولوجية مهيئة لذلك فلا يجوز مثلاً أن تكون مساحة المربع ١ سم<sup>٢</sup> فقط فهذه المساحة أصغر من جعل الطحالب تتوزع بواسونياً . كذلك في تسجيل حالات الانفلاتر التي ترد إلى مستشفى كبير لا ينبغي أن تقل وحدة الزمن عن أسبوع مثلاً ، لإعطاء الفرصة لورود حالات كافية .

( ثانياً ) استقلال الأحداث ( عشوائية وقوع الأحداث ) :

يشترط أن يكون وقوع الأحداث عشوائياً بمعنى أن يكون احتمال وقوع أو عدم وقوع الحدث في أى لحظة مستقلاً عن وقوعه أو عدم وقوعه في أى لحظة سابقة أو لاحقه غير متداخلة معها وبذلك لا يتأثر وقوع الحدث إلا بالعوامل العشوائية وحدها . فمثلاً وجود طحلب في جزء من مربع ما لا ينبغي أن يزيد أو ينقص من احتمال نمو طحالب أخرى في أى جزء آخر من المربع . كذلك تسجيل حالة انفلاتر في لحظة ما لا يجب أن يؤثر في احتمال تسجيل حالات تالية .

تحت هذين الشرطين اتضح أنه بتقريب جيد إلى حد كبير أو بالضبط يمكن إيجاد احتمال عدد مرات وقوع الحدث في أى فترة زمنية أو فراغية بواسطة توزيع بواسون دليhle :

(١١)

٢ = e<sup>-س</sup>

أى من الدالة د (س) =  $\frac{e^{-س} س^ك}{ك!}$  حيث س = ٠ ، ١ ، ٢ ، ....

وحيث د (س) هو احتمال وقوع الحدث س من المرات خلال فترة زمنية أو فراغية ز وحيث ك مقدار ثابت موجب يعبر عن متوسط وقوع الحدث في وحدة الزمن أو الفراغ ، وهذا المقدار يحسب تخريبياً .

مثال (٣ - ٧) :

إذا فرض أن البكتريا من نوع معين تتواجد في الماء بمعدل ٢ بكتريا في السنتيمتر المكعب وأن عدد البكتريا يتوزع توزيعاً بواسونياً فأوجد احتمال أنه في عينة عشوائية من ٣ سم<sup>٣</sup> من الماء : (أ) لا يوجد بكتريا (ب) يوجد ٣ بكتريا على الأقل .

الحل :

لدينا ك = ٢ بكتريا في السنتيمتر المكعب ، ز = ٢ سم<sup>٣</sup> .

إذن م = ٢ × ٢ = ٤

د (س) =  $\frac{e^{-m} m^s}{s!}$  حيث س = ٠ ، ١ ، ٢ ، ...

(أ) احتمال عدم وجود بكتريا في ٢ سم<sup>٣</sup> د = (٠) =  $e^{-4} = ٠,١٨٣٢$

(ب) احتمال وجود ٣ بكتريا على الأقل في ٢ سم<sup>٣</sup> ل (س ≤ ٣) .

١ - ١ = ل (س ≥ ٢) = ١ - [د (٠) + د (١) + د (٢)]

١ - ١ = (١ + ٤ + ٨)  $e^{-4} = ١ - ١٣ e^{-4}$  .

٠,٧٦١٨٤ =

مثال (٣ - ٨) :

إذا كان متوسط عدد السيكلوب في لتر من ماء بحيرة هو ٢ فما احتمال وجود

٥ سيكلوب على الأقل في عينة من ٣ لترات من ماء البحيرة ؟

( السيكلوب كائن دقيق يطفو على الماء وتقتات عليه الأسماك ) .

الحل :

لدينا ك = ٢ سيكلوب في اللتر ، م = ٣ لتر



$$\text{إذن } ٢ = ٣ \times ٢ = ٦$$

$$٦, د (س) = \frac{٦}{١!} هـ ٦ حيث س = ٠, ١, ٢, \dots$$

احتمال وجود ٥ سيكلوب على الأقل = ل (س ≤ ٥) = ١ - ل (س ≥ ٤)

$$= ١ - ٠,٢٨٥$$

$$= ٠,٧١٥$$

(٣ - ٤ - ٣) اختبار استقلال الأحداث النادرة (أو اختبار العشوائية) :

عند تناول حدث نادر يهتما في كثير من الأحيان دراسة استقلال الأحداث أى دراسة ما إذا كان وقوع الحدث في لحظة ما يزيد أو ينقص من احتمال وقوعه في لحظة تالية كما هو الحال مثلا في دراسة توزيع سوسة الفاصوليا أو توزيع حشرة على نوع من الذباب . ونظراً لأن الحدث النادر لا يتوزع بواسونيا إلا إذا كانت الأحداث مستقلة فإننا نستطيع الحكم على استقلال الأحداث عن طريق اختبار ما إذا كان التوزيع بواسونيا ، وهذا الاختبار يمكن إجراؤه بتوفيق توزيع بواسون لتوزيع تكرارى مشاهد في تجربة كما فعلنا في حالة ذى الحدين في البند (٢ - ٣ - ١) والمثال (٣ - ٤) . فإذا كانت المطابقة حسنة دل ذلك على أن التوزيع بواسونيا وبالتالي تكون الأحداث مستقلة وتقع عشوائياً ، أما إذا لم تكن المطابقة حسنة فنحكم بعدم عشوائية وقوع الأحداث .

مثال (٣ - ٩) :

أجريت تجربة لاختبار توزيع خلايا الخميرة في ٤٠٠ مربع من جهاز الـ hemacytometer (صندوق لعد الخلايا) ووجد التوزيع التكرارى المبين بالعمودين الأول والثانى من الجدول (٣ - ٢) . بالتأمل في هذين العمودين نلاحظ أمرين هامين هما :

(أ) أن ٧٥ من هذه المربعات أى حوالى ١٩٪ منها لا تشتمل على أى خلية ومعظم المربعات (٥٦٪) تحمل إما خلية واحدة أو خليتين وأن ١٧ مربعاً فقط (أى حوالى ٤٪) تحتوى على ٥ خلايا أو أكثر.

(ب) الوسط الحسابى  $\bar{x} = \frac{1}{n} \sum x_i$

$$\frac{1}{400} (9 \times 1 + \dots + 2 \times 121 + 1 \times 103 + 0 \times 75) = 1,8 = \frac{720}{400}$$

الجدول (٣ - ٢)

التكرارات المشاهدة والتكرارات البواسونية المتوقعة لعدد خلايا الحميرة في ٤٠٠ مربع من جهاز الهامستر

عدد الخلايا في المربع س	التكرارات المشاهدة ك	النسبة المتوقعة ل = د = (س) ق = ط × ٤٠٠	التكرارات المتوقعة	الانحرافات ك - ق
٠	٧٥	٠,١٦٥٣	٦٦,١	+
١	١٠٣	٠,٢٩٧٥	١١٩,٠	-
٢	١٢١	٠,٢٦٧٨	١٠٧,١	+
٣	٥٤	٠,١٦٠٧	٦٤,٣	-
٤	٣٠	٠,٠٧٢٣	٢٨,٩	+
٥	١٣	٠,٠٢٦٠	١٠,٤	+
٦	٢	٠,٠٠٧٨	٣,١	-
٧	١	٠,٠٠٢٠	٠,٨	+
٨	٠	٠,٠٠٠٤	٠,٢	-
٩	١	٠,٠٠٠١	٠,٠٤	-
	٤٠٠	٠,٩٩٩٩	٣٩٩,٩	

أى أن متوسط عدد الخلايا في المربع هو ١,٨ خلية وهذا عدد صغير فعلا بالنسبة لسعة كل مربع وبالنسبة لعدد الخلايا التي يحتمل أن تظهر في أى من المربعات .

من هذا يحق لنا أن نعتبر أن الحدث هو حدث نادر ونتوقع بالتالى أن يكون توزيعه بواسونيا إذا توفر شرط الاستقلال . ولاختبار هذا الشرط نوفق توزيع بواسون للتوزيع التكرارى المشاهد مع تقدير الدليل م لذلك التوزيع من العينة أى نأخذ  $m = 1,8$  فكون دالة كتملة الاحتمال :

$$D(s) = \frac{1,8^s}{s!} e^{-1,8} \text{ حيث } s = 0, 1, 2, \dots$$

نحسب الاحتمالات  $D(0), D(1), D(2), \dots$  كما في العمود الثالث من الجدول (٣ - ٢) ثم نضرب كلا من هذه الاحتمالات ( التكرارات النسبية المتوقعة ) في حجم التوزيع التكرارى المشاهد وهو ٤٠٠ فنحصل على التكرارات المتوقعة المبينة بالعمود الرابع .

بمقارنة التكرارات المشاهدة بالتكرارات النظرية نجد أن هناك مطابقة حسنة بين التوزيع التكرارى المشاهد وتوزيع بواسون دليله ١,٨ ولا يوجد غلط واضح لانحرافات التكرارات المشاهدة عن التكرارات المتوقعة لها كما يبدو من العمود الخامس وإن كان الحكم الموضوعى لهذا التطابق لا يتأق إلا بأحد الاختبارات الإحصائية التي سندرسها بعد . انظر المسألة (٦) في تمارين (٦ - ٢) .

ونستنتج من هذا أن توزيع خلايا الخميرة هو توزيع بواسوني وهذا يتضمن أن الأحداث هنا تقع عشوائياً أى مستقلة عن بعضها .

وهناك اختبار آخر يساعد على بيان ما إذا كان التوزيع بواسونياً دون الالتجاء إلى عطية التوفيق ، ويعتمد هذا الاختبار على الخاصة الهامة التي جاءت في المساوية (٩) عن تماوى التباين والوسط الحسابي في التوزيعات البواسونية . ونحن نتناول

عينة عشوائية من مجتمع بواسوني تتوقع وجود هذا التساوى بالتقريب أى نتوقع أن تكون النسبة بين التباين والوسط الحسابي المحسوين من العينة قريبة من الواحد الصحيح . في المثال الأخير نجد أن :

$$\text{تباين العينة ع}^2 = \frac{1}{n} \left[ \sum_{i=1}^n (x_i - \bar{x})^2 \right] = \frac{1}{399} [720 - (81 \times 1 + \dots + 4 \times 121 + 1 \times 103 + 0 \times 75)] =$$

$$1,965 =$$

وهذا العدد قريب من الوسط الحسابي 1,8 كما أن النسبة بينهما  $\frac{1,965}{1,8} = 1,092$

قريبة من الواحد وهذا يدعم استنتاجنا السابق بأن توزيع المتغير هو توزيع بواسوني وما يتبع ذلك من عشوائية الأحداث . هذا ويمكن اختيار كبير أو صغر النسبة ع' / س من الواحد بطريقة موضوعية عن طريق اختبارات الذى سندرسه بعد . انظر المثال (6 - 5) بالبند (6 - 6) .

**نمط التجمع ونمط التناثر :**

إذا اتضح أن التكرارات المشاهدة تنحرف عن التكرارات المتوقعة لها بشكل جوهري أو أن النسبة ع' / س أكبر أو أصغر من الواحد بشكل جوهري فإن هذا يعني أن الأحداث لا تقع مستقلة عن بعضها بل يؤثر وقوع أو عدم وقوع أحدها في وقوع أو عدم وقوع الأحداث الأخرى . وهذا يدعو إلى التساؤل عما إذا كانت التكرارات المشاهدة تنم عن نمط خاص وعن تفسير ما قد يوجد من أنماط . وهذا التفسير لا يستطيع التحليل الإحصائي وحده القيام به فالمرجع الأول في هذا هو معرفة الباحث بظروف التجربة وطبيعة المتغيرات والعوامل التي تؤثر فيها .

وهناك نوعان رئيسيان من الأنماط هما :

### CLUMPING

(١) نمط التجمع :

في هذا النمط تكون التكرارات المشاهدة أكبر من التكرارات المتوقعة عند ذيلي التوزيع وأصغر منها عند الوسط كما هو الحال في التجربة الملخصة بالجدول (٣ - ٣) الآتي الذي يعرض توزيع الحلم المائي water mite على ٥٨٩ هاموشة . ويوصف هذا النمط بأنه مُعْدَى contagious بمعنى أن وقوع حدث ( ظهور حلمة مثلا ) يرفع من احتمال وقوع أحداث أخرى والعكس بالعكس . وفي هذه الحالة تكون النسبة بين التباين والوسط الحسابي في المجتمع أكبر من الواحد .

### REPULSION

(٢) نمط التناثر :

في هذا النمط تكون التكرارات المشاهدة أصغر من التكرارات المتوقعة عند الدليلين وأكبر منها عند الوسط كما هو الحال في التجربة الملخصة بالجدول (٣ - ٤) الذي يسجل توزيع السوس على ١١٢ نبات فاصوليا . وهنا يكون وقوع الحدث ( خروج سوسة مثلا ) عائقاً لوقوع أحداث أخرى فيشتمل التوزيع على عدد قليل من المجموعات المتجانسة وعدد كبير من المجموعات المختلطة . وفي هذه الحالة تكون النسبة بين التباين والوسط الحسابي في المجتمع أصغر من الواحد .

بعد حساب الوسط الحسابي والتباين للتوزيع المشاهد نجد ما يلي :

$$\text{بالنسبة للتجربة الأولى : } \bar{x} = 0.9709 / 0.4343 = 2.236$$

هذه النسبة أكبر جوهرياً من الواحد . انظر البند (٦ - ٦ - ٤) .

$$\text{بالنسبة للتجربة الثانية : } \bar{x} = 0.269 / 0.4643 = 0.579$$

هذه النسبة أصغر جوهرياً من الواحد . انظر البند (٦ - ٦ - ٤) .

الجدول (٣-٣)  
التكرارات للشاذلة والتكرارات الواسونية للوقعة  
لعدد الحظم على ٥٨٩ هاموشه (نط نجح)

عدد الحظم على الهاموشة	ك	ق	ك - ق
٠	٤٤٢	٣٨٠,٧	+
١	٩١	١٦٦,١	-
٢	٢٩	٣٦,٢	-
٣	١٤	٥,٣	+ {
٤	٤	٠,٦	+ {
٥	٦	٠,١	+ {
٦	٢	٠,٠	+ {
٧	٠	٠,٠	٠
٨	١	٠,٠	+ {
	٥٨٩	٥٨٩	

الجدول (٤-٣)  
التكرارات للشاذلة والتكرارات الواسونية للوقعة  
لعدد السوس الذى خرج من ١١٢ نبات لاصوليا (نط تالف)

عدد السوس على النبات	ك	ق	ك - ق
٠	٦١	٧٠,٤	-
١	٥٠	٣٢,٧	+
٢	١	٧,٦	- {
٣	٠	٨,٩١,٢	- {
٤	٩	٠,١	- {
	١١٢	١١٢	

## تمارين (٣ - ٢)

(١) في توزيع بواسون دليله  $\mu = ٠,٧٢$  أوجد :  
 ل (س = ٠) ، ل (س = ١) ، ل (س = ٢) ، ل (س < ٢) .  
 ( اعتبر أن  $٧٢ = ٠,٤٨٦٨$  )

(٢) دلت الخبرة الطويلة على أن السفن تدخل في إحدى المواني بمعدل ٣ سفن في الساعة . إذا كان توزيع عدد السفن التي تدخل هذا الميناء هو توزيع بواسون فأوجد احتمال أنه في فترة زمنية قدرها دقيقتان ( أ ) لا تدخل أى سفينة (ب) تدخل سفينتان على الأقل .

(٣) إذا علم أن الجسيمات تنبعث من مصدر مشع بمعدل ٠,٥ جسيماً في الثانية وأن عدد هذه الجسيمات يتوزع بواسونياً فاحسب احتمال انبعاث ٣ جسيمات أو أكثر في فترة زمنية طولها ٦ ثواني .

(٤) الجدول الآتي يعرض توزيع عدد الحلم ( المايث المائية ) water mite على نوع من الذباب . وفق توزيعاً بواسونياً لهذا التوزيع واستنتج أن وقوع الأحداث ( ظهور الحشرات على الذبابة ) ليس عشوائياً . أوجد تباين التوزيع التكرارى المشاهد وقارنه بالوسط الحسابي لتدعيم استنتاجك ( ستجد أن النسبة بين التباين والوسط الحسابي ٢,٢٢٥ ) .

عدد الحشرات على الذبابة ٠ ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ المجموع  
 ٤٢٢ ٩١ ٢٩ ١٤ ٤ ٢ ١ ٠ ٥٨٩ التكرارات المشاهدة

(٥) أجب عن نفس السؤال السابق مستخدماً التوزيع التكرارى الآتي الذى يعرض توزيع عدد السوس على قرن فاصوليا ( قيس هذا العدد بعدد الثقوب التي نتجت في قرن الفاصوليا أثناء خروج اليرقات منها ) .

عدد السوس في القرن : ٠ ١ ٢ ٣ ٤ المجموع  
 ٦١ ٥٠ ١ ٠ ٠ ١١٢ التكرارات المشاهدة

(٦) يدخل الزبائن في أحد المحلات بمعدل ٣٠ شخصاً في الساعة فإذا كان توزيع عدد الزبائن هو توزيع بواسوني فأوجد احتمال أنه في فترة زمنية قدرها دقيقتان (أ) لا يدخل أحد (ب) يدخل شخصان على الأقل .

### (٣ - ٥) توزيع باسكال : PASCAL DISTRIBUTION

في توزيع ذي الحدين نعتبر أن حجم العينة هو عدد ثابت  $n$  ونعالج متغيراً عشوائياً  $x$  يعبر عن عدد مرات وقوع الحدث في  $n$  من المحاولات ، إلا أن هناك مشكلات تتطلب عكس هذا الوضع فيكون حجم العينة متغيراً عشوائياً  $x$  ويكون عدد مرات وقوع الحدث هو عدد ثابت  $a$  محدد من قبل ويكون المطلوب هو إيجاد احتمالات قيم المتغير  $x$  التي تسمح بوقوع الحدث هذا العدد المحدد من المرات . ( ملاحظ أن المعاينة هنا تكون من النوع التتابعي sequential sampling ) .  
أى أننا نعالج متغيراً عشوائياً وثابتاً  $x$  يعبر عن العدد اللازم من المحاولات لكي يقع الحدث عدداً محدداً  $a$  من المرات .

وبطبيعة الحال لا يجب أن يقل عدد المحاولات عن العدد  $a$  لأن وقوع الحدث  $a$  من المرات يتطلب  $a$  محاولة على الأقل ، وعلى ذلك تبدأ قيم  $x$  بالعدد  $a$  . أى أن هذا المتغير لا يأخذ إلا القيم  $a$  ،  $a+1$  ،  $a+2$  ، ... فإذا كان المطلوب وقوع الحدث  $a$  مرات مثلاً فإن  $x$  تأخذ القيم  $a$  ،  $a+1$  ،  $a+2$  ، ...

تحت نفس افتراضات العشوائية وثبات الدليل  $a$  واستقلال الأحداث يمكن أن نثبت رياضياً أن دالة كثرة الاحتمال لهذا المتغير تأخذ الصورة الآتية :

$$P(x) = \frac{e^{-\lambda} \lambda^a}{a!} \quad \text{حيث } \lambda = a + 1, a + 2, \dots \quad (١٢)$$

وحيث  $0 < \lambda < 1$  ،  $\lambda = a - 1$  ،  $a$  مقدار ثابت



ويسمى توزيع الاحتمال حيثخذ بتوزيع باسكال دليلاه ح ، أ كما يسمى المتغير  $\bar{x}$  بمتغير باسكال . يلاحظ أن الحدث يقع أ من المرات المستقلة باحتمال ح في كل مرة وأنه يفشل في س - أ من المرات المستقلة باحتمال ك في كل مرة ، وأن المرة الأخيرة هي نجاح دائما . كما يلاحظ أن عدد الطرق التي تقع فيها ( أ - ١ ) نجاحات من ( س - ١ ) محاولات هو  $\binom{1-s}{1}$  وعلى ذلك فإن احتمال وقوع الحدث أ من المرات ( وفشله س - أ من المرات ) هو  $\binom{1-s}{1} \cdot \bar{c}^1 \cdot c^{s-1}$

مثال ( ٣ - ١٥ ) :

المعروف أن ٦٠٪ من المرضى بمرض معين يستجيبون للدواء ما بعد تناوله لمدة أسبوع . يختار مريض بهذا المرض الواحد بعد الآخر في ترتيب عشوائي ليتناولوا الدواء ( لمدة أسبوع ) حتى يحصل على ٥ استجابات صحيحة :

( أ ) ما احتمال أن يكون العدد المختار من المرضى سبعة ؟

( ب ) ما احتمال أن يكون العدد المختار من المرضى عشرة ؟

الحل :

إن عدد المرضى اللازمين للحصول على خمس استجابات صحيحة هو متغير عشوائي  $\bar{x}$  له توزيع باسكال دليلاه ٠,٦ ، ٥ ودالة كتلة احتماله :

$$P(\bar{x}) = \binom{1-s}{x} \cdot \bar{c}^x \cdot c^{s-x} \quad \text{حيث } s = 5, 6, 7, \dots$$

$$(أ) \quad P(\bar{x}=5) = P(5) = \binom{1-s}{5} \cdot \bar{c}^5 \cdot c^{s-5}$$

$$= 0,1866$$

$$(ب) \quad P(\bar{x}=6) = P(6) = \binom{1-s}{6} \cdot \bar{c}^6 \cdot c^{s-6}$$

$$= 0,1003$$

## ملاحظة :

إن الحالات التى يظهر فيها متغير باسكالى تنشأ فى المعتاد عندما يستخدم ما يسمى بالمعانة المتتابعة sequential sampling حيث لا يحدد حجم العينة مسبقاً ، بل تختار المشاهدات فى تتابع عشوائى الواحدة بعد الأخرى وتتوقف هذه العملية حين يتجمع عدد كاف من المشاهدات يمكننا من اتخاذ قرار بحسب قاعدة معينة توضع سلفاً . ففى المثال (٣ - ١٠) نفرض أن المطلوب اختبار صحة الفرض أن ٦٠٪ من المرضى يستجيبون للدواء ، وأن القاعدة التى وضعت لتحديد صحة أو خطأ هذا الفرض كالآتى :

« اختر المرضى الواحد بعد الآخر بترتيب عشوائى ليتناول الدواء وسجل العدد  $s$  لعدد المرضى المختبرين حتى نحصل على ٥ استجابات صحيحة . ارفض الفرض إذا كان الاحتمال  $L(s \leq s_0)$  يساوى أو يقل عن ٠,٠٥ ، وإلا فاقبل الفرض » .

إذا استخدمت هذه القاعدة فماذا يكون حكمنا عن الفرض إذا وجد فى تجربة ما أن عدد المرضى المختبرين حتى الوصول إلى ٥ استجابات صحيحة هو :

$$(أولاً) : s_0 = 8 \quad (ثانياً) : s_0 = 13$$

الحل :

$$\begin{aligned} (أولاً) \quad L(s \leq 8) = 1 - [D(7) + D(6) + D(5)] - 1 \\ = [C(0,4) \times C(0,6)^4 \times 0,4 + C(0,6)^4 \times 0,4] - 1 = \\ = 0,58 = [C(0,6)^4 \times (1 + 2 + 4)] - 1 \\ \text{بما أن } 0,58 > 0,05 \text{ نقبل الفرض أن نسبة الشفاء } 60\% \end{aligned}$$

$$\begin{aligned} (ثانياً) \quad L(s \leq 13) = 1 - [D(12) + D(11) + \dots + D(6) + D(5)] - 1 \\ = [C(0,6)^4 \times (1 + 2 + 4 + \dots + 12 + 13)] - 1 = \\ = 0,03 < 0,05 \text{ نرفض الفرض أن نسبة الشفاء } 60\% . \end{aligned}$$

### (٣ - ٦) التوزيع الهندسي : THE GEOMETRIC DISTRIBUTION

هو حالة خاصة من توزيع باسكال تكون فيها عدد مرات وقوع الحدث  $أ = ١$  ويرمز المتغير  $س$  هنا إلى عدد المحاولات اللازمة لوقوع الحدث لأول مرة .  
وتنتج دالة كتلة الاحتمال لهذا المتغير بوضع  $أ = ١$  في الصيغة (١٢) أى تكون على الصورة :

$$د(س) = ح ك^{س-١} \quad \text{حيث } س = ١, ٢, ٣, \dots (١٣)$$

ويسمى التوزيع حيثث بالتوزيع الهندسي أو بتوزيع وقت الانتظار waiting time distribution ولهذا التوزيع دليل واحد هو ح . وهو يفيد في دراسة الخواص النادرة للمجتمعات كما هو الحال في أمراض الدم النادرة حيث قد لا ينفعنا استخدام توزيع ذى الحدين ، لأننا لو حددنا حجم العينة فقد لا يقع الحدث في أى عنصر منها وبذلك لا نحصل على معلومات كافية عن الحدث ، بينما التوزيع الهندسي يضمن وقوع الحدث .

يمكن أن تثبت رياضياً أن :

$$\frac{١}{ح} = \text{الوسط الحسابي للتوزيع} \quad (١٤)$$

$$\frac{ك}{١-ك} = \text{تباين التوزيع} \quad (١٥)$$

مثال (٣ - ١١) :

في إحدى المنتجات الصناعية المعروف أنه في المتوسط توجد وحدة معيبة في كل ١٠٠ وحدة . تختار وحدات عشوائياً وتعتبر الواحدة بعد الأخرى إلى أن تظهر أول وحدة معيبة . ( أ ) ما احتمال اختبار ٥ وحدات حتي الوصول إلى الوحدة المعيبة ؟ (ب) ما العدد المتوقع للاختبارات اللازمة للعثور على أول وحدة معيبة ؟

الحل :

لدينا توزيع هندسي دليله  $0.01 = 0.99$  وإذن دالة كثة احتمال  
 د (س)  $= \sum_{k=0}^{\infty} 0.01 (0.99)^k$  حيث  $s = 1, 2, 3, \dots$

$$(أ) د (0) = 0.01 (0.99)^0$$

$$= 0.0099$$

(ب) العدد المتوقع يعني الوسط الحسابي  $= \frac{1}{0.01} = \frac{1}{0.01} = 100$  اختبار .

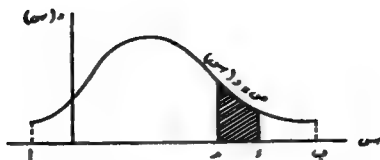
### (٣ - ٧) توزيعات الاحتمال المتصلة :

تمتد الأفكار السابقة عن توزيعات الاحتمال للمتغيرات الوثابة إلى حالة المتغيرات المتصلة. مع بعض الفروق التي تقتضيها طبيعة كل من هذين النوعين من المتغيرات . فإذا كان  $x$  متغيراً حقيقياً من النوع المتصل مداه الفترة ( أ ، ب ) فإن احتمال أن يأخذ هذا المتغير قيمة محددة  $s$  يساوى صفراً ، ذلك لأن أى متغير متصل يأخذ عدداً لا نهائياً من القيم في أى جزء من مداه مهما كان صغيراً .

ولذلك يعرف توزيع الاحتمال في هذه الحالة بواسطة دالة متصلة غير سالبة د حيث :

$$L(s) = \Delta(s) = \int_{-\infty}^{+\infty} \delta(s-s') ds' \quad (16)$$

وهذه الصيغة تعني أن احتمال وقوع قيم المتغير  $x$  في فترة  $\Delta$   $s$  يساوى تكامل الدالة د على هذه الفترة . إن مثل هذه الدالة تسمى بدالة كثافة الاحتمال probability density function وتمثل بياناً بالشكل (٣ - ٣) .



الشكل (٣-٣) توزيع الاحتمال لمعدل

ومن الخواص الرئيسية للمنحنى الممثل لأى دالة كثافة احتمال أن مساحة المنطقة الواقعة أسفله وفوق المحور السيني تساوى الواحد الصحيح ، ويمكن رؤية هذا المنحنى كخط ممهد لمضلع تكرارى يمثل التكرارات النسبية لتوزيع تكرارى ذى فئات مبني على عدد كبير جداً من المشاهدات موزعة على عدد كبير جداً من الفئات ذوات الأطوال الصغيرة جداً .

ومن تعريف الدالة د نرى أن احتمال وقوع قيم المتغير س بين عددين ج ، د أى في فترة ( ج ، د ) محتواة في المدى ( أ ، ب ) يعطى بالتكامل .

$$P(J < S < D) = \int_J^D d(S) \cdot dS \quad (١٧)$$

أى أننا إذا اخترنا عشوائياً قيمة واحدة من قيم المتغير فإن احتمال وقوعها بين عددين ج ، د يعطى بالتكامل المذكور . ومن الواضح أن هذا التكامل يساوى عددياً مساحة الجزء المظلل بالشكل ( ٣ - ٣ ) ، كما أنه يعبر عن نسبة قيم المتغير الواقعة بين عددين ج ، د .

ويعرف الوسط الحسابي  $\mu$  والتباين  $\sigma^2$  لتوزيع احتمال متغير متصل مداه الفترة  
(أ، ب) كالآتي :

$$(18) \quad \int_a^b x \cdot dF(x) = \mu$$

$$(19) \quad \int_a^b (x - \mu)^2 \cdot dF(x) = \sigma^2$$

$$\text{أو} \quad \int_a^b x^2 \cdot dF(x) = \mu^2 + \sigma^2$$

مثال (٣ - ١٢) :

أوجد الوسط الحسابي والتباين لتوزيع احتمال المتغير المتصل الذى دالة كثافة  
احتماله هي :

$$dF(x) = (x-1) \cdot dx \quad \text{حيث } 1 < x < 2.$$

الحل :

$$\int_1^2 x \cdot (x-1) \cdot dx = \mu$$

$$\int_1^2 (x^2 - x) \cdot dx = \mu$$

$$\left[ \frac{x^3}{3} - \frac{x^2}{2} \right]_1^2 = \mu$$

$$\left( \frac{8}{3} - 2 \right) - \left( \frac{1}{3} - \frac{1}{2} \right) = \mu$$

$$\frac{2}{3} - \left( -\frac{1}{6} \right) = \mu$$

$$\frac{2}{3} + \frac{1}{6} = \mu$$

$$\frac{4}{6} + \frac{1}{6} = \mu$$

$$\frac{5}{6} = \mu$$

ومن أشهر توزيعات الاحتمال المتصلة وأكثرها استخداماً ذلك التوزيع المسمى بالتوزيع المعتدل الذى يلعب دوراً رئيسياً في النظرية الإحصائية ونظرية الاحتمال كما أن هناك توزيعات احتمالات متصلة أخرى تستخدم كنماذج احتمال لبعض المجتمعات ، منها التوزيع المعتدل اللوغاريتمى والتوزيع المعتدل الدائرى وتوزيع جاما والتوزيع الأسى ... وسوف نكتفى بتقديم التوزيع المعتدل والتوزيع المعتدل اللوغاريتمى في الفصل القادم .





## الفصل الرابع

### التوزيع المعتدل والتوزيع اللوغاريتمى

#### THE NORMAL DISTRIBUTION AND THE LOGNORMAL DISTRIBUTION

##### ( أولا ) التوزيع المعتدل

التوزيع المعتدل هو أهم توزيعات الاحتمال القياسية وأكثرها استخداما ، إذ يؤخذ كنموذج لكثير من المتغيرات المتصلة ، منها أطوال بتلات بعض النباتات - أطوال أجنحة الذباب المنزل - أطوال وموازن الأطفال عند الولادة - مقادير الدسم فى الزبد الناتج من بعض أنواع الأبقار ... فضلا عن أنه يلعب دورا كبيرا فى بناء نظرية الاحتمال والنظرية الإحصائية .

وقد سُمى هذا التوزيع بالتوزيع المعتدل ( أو المعتاد أو الطبيعى ) لأنه كان يظن فيما مضى أن أية بيانات عن ظواهر الحياة ينبغى أن تمثل لهذا التوزيع وإلا كانت هذه البيانات مشكوكا فيها . إلا أنه ثبت الآن أن الأمر ليس كذلك فهناك كثير من الظواهر ذات توزيعات تختلف عن التوزيع المعتدل . كما يسمى التوزيع بتوزيع جاوس اعترافا بفضل العالم الألماني كارل فردريك جاوس ( ١٧٧٧ - ١٨٥٥ ) لذى استنبط التوزيع رياضيا كتوزيع احتمال أخطاء القياس وكان لذلك يسميه

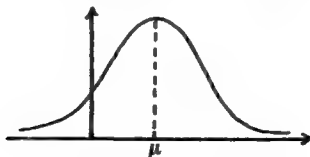
بالقانون الطبيعي للأخطاء normal law of errors ، ويسمى أيضا بتوزيع جاوس - لابلاس اعترافا بجميل العالم الفرنسي بيير سيمون لابلاس (١٧٤٩ - ١٨٢٧) الذى كان أول من اكتشفه وأثبت صلاحيته كنموذج لكثير من الظواهر .  
ويعرف هذا التوزيع بواسطة دالة كثافة الاحتمال المعطاة بالقاعدة :

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{حيث } -\infty < x < \infty , 0 < \sigma < \infty \quad (١)$$

ولهذه الدالة بارامتران هما  $\mu$  ،  $\sigma$  ، تعبر  $\mu$  عن الوسط الحسابى وتعبر  $\sigma$  عن الانحراف المياري للتوزيع . ويتحدد التوزيع تماما إذا علمت قيمتا هذين الدليلين ، ولذا نرمز له بالرمز مع  $(\mu, \sigma)$  . والمتغير العشوائى  $x$  الذى له هذه الدالة يسمى بالمتغير المعتدل .

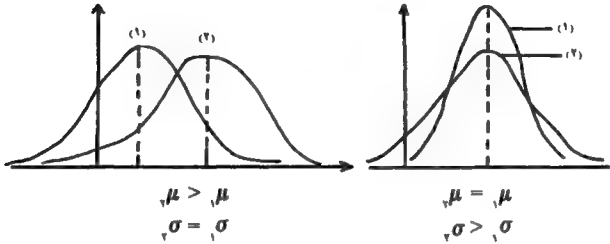
#### (٤ - ١) بعض خواص التوزيع

(أ) المنحنى الذى يمثل الدالة (١) يسمى بالمنحنى المعتدل وهو منحنى ذو قمة واحدة ومتماثل حول الخط  $x = \mu$  . ومساحة المنطقة الواقعة بين هذا المنحنى ومحور السينات تساوى الواحد الصحيح كما هو الحال لمنحنى أى توزيع احتمال . ومن التماثل نجد أن الخط  $x = \mu$  يقسم الشكل إلى منطقتين متساويتى المساحة ومساحة كل منهما تساوى  $\frac{1}{2}$  . انظر الشكل (٤ - ١) .



الشكل (٤ - ١) : منحنى التوزيع المعتدل

(ب) لكل من البارامترين  $\mu$  ،  $\sigma$  عدد غير منتهى من القيم ، ولذلك هناك عدد غير منتهى من التوزيعات المعتدلة . هذا مع ملاحظة أن  $\mu$  هو بارامتر موضع location parameter أما  $\sigma$  فهو بارامتر شكل shape parameter كما يتبين من الشكل (٤ - ٢) الآتي :



الشكل (٤ - ٢)

(ج) معامل الالتواء = صفر ، معامل التفرطح = ٣

(د) إذا كانت  $s$  ترمز إلى قيم متغير معتدل  $\bar{x}$  وسطه الحسابي  $\mu$  وانحرافه المعياري  $\sigma$  فإن الصيغة :

$$ع = \frac{\mu - s}{\sigma} \quad (٢)$$

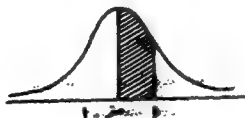
تحول المتغير  $\bar{x}$  إلى متغير  $ع$  له توزيع معتدل وسطه الحسابي صفر وانحرافه المعياري الواحد الصحيح ، ويسمى هذا المتغير بالمتغير المعتدل المعياري وسنرمز له بالرمز : مع (١ ، ٠) .

(هـ) إذا رسم توزيع التكرارات النسبية للمتجمعة لمتغير معتدل على ورق الرسم البياني ذى التقسيم الخطي ( العادي ) فإن هذا المنحني يتخذ الشكل  $s$  غير أن

هناك ورق يسمى ورق تقسيم الاحتمالات المعتدلة normal probability graph paper إذا رسم عليه التوزيع نتج خط مستقيم ، ويستخدم هذا الورق للكشف عن اعتدالية المجتمعات كما سنرى في البند (٤ - ٤) الآتي .

#### (٤ - ٢) جداول المساحات أسفل المنحني المعتدل المعيارى :

نظراً لأهمية معرفة احتمالات المتغيرات المعتدلة وكثرة الحاجة إليها فقد حسبت قيم هذه الاحتمالات لمختلف قترات المتغير المعتدل المعيارى مع ووضعت في جداول تعرف بـ جداول المساحات أسفل المنحني المعتدل المعيارى ، وهذه الجداول تأخذ صوراً متعددة تؤدي إلى نفس النتائج ومنها الصورة الواردة بالجدول (٦) بملحق هذا الكتاب . وهذا الجدول يعطى للمساحة تحت المنحني المعتدل المعيارى وفوق الفترة بين الصفر وبعض أعداد موجبة أ متوقعة للمتغير مع ، وهذه هي المساحة المظلمة بالشكل (٤ - ٣) وهي تعبر عن الاحتمال ل (  $0 < z \leq 1$  ) (٣)



الشكل (٤ - ٣)

أى عن احتمال وقوع قيم المتغير مع بين العددين ، أى فى الفترة ( . ، أ ) . ويلاحظ من تماثل المنحني أن هذا الاحتمال هو أيضاً احتمال وقوع قيم المتغير مع بين العددين - أ ، . أى : ل (  $0 \leq z < 1$  ) (٤)

وذلك لتساوى مساحتي المنطقتين المناظرتين .

في هذا الجدول يعبر الهامش الرأسي ( الذى على اليسار أو على اليمين ) عن

قيم أ إلى خانة عشرية واحدة ، ويعبر الفامش الأقعي ( الذى في أعلى الجدول )  
عن الخانة العشرية الثانية أى خانة الجزء من مائة . أما الأعداد التي في قلب الجدول  
فهى احتمالات وقوع المتغير ع في الفترة ( ٠ ، أ ) .

مثال ( ٤ - ١ ) :

إذا علم أن توزيع درجات الطلاب في مادة ما هو توزيع معتدل وسطه الحسابي  
 $\mu = ٦٠$  وانحرافه المعياري  $\sigma = ١٦$  . فأوجد باستخدام جدول المساحات  
الاحتمالات الآتية :

أولاً : ل (  $٨٠ \leq س < ٦٠$  )      ثانياً : ل (  $س < ٨٠$  )  
ثالثاً : ل (  $س \geq ٨٠$  )      رابعاً : ل (  $٧٦ \leq س < ٤٠$  )  
خامساً : ل (  $٨٥ \leq س < ٧٥$  )

الحل :

لإيجاد الاحتمالات المطلوبة من الجدول يجب أن نحول المتغير المعتدل س إلى المتغير  
المعتدل المعيارى ع بواسطة الصيغة (٢) وهى هنا

$$ع = \frac{س - ٦٠}{١٦}$$



$$\text{(أولاً) بوضع } س = ٦٠ \text{ نجد أن } ع = \frac{٦٠ - ٦٠}{١٦} = \text{صفر}$$

$$\text{وبوضع } س = ٨٠ \text{ نجد أن } ع = \frac{٦٠ - ٨٠}{١٦} = ١,٢٥$$

$$\text{إذن ل ( } ٦٠ \leq س < ٨٠ \text{ ) = ل ( } ٠ < ع \leq ١,٢٥ \text{ )}$$

$$= ٠,٣٩٤٤$$

وذلك من الجدول مباشرة عند العدد ١,٢ الذى فى الهامش الرأسى وتحت العدد ٥ الذى فى الهامش الأفقى . وهذه النتيجة تعنى أن حوالى ٣٩٪ من الطلاب تقع درجاتهم بين ٦٠ ، ٨٠ .



( ثانياً ) بوضع  $s = ٨٠$  نجد أن  $ع = ١,٢٥$

$\therefore (s < ٨٠) ل = (١,٢٥ < ع)$

= مساحة المنطقة التى على يمين العدد ١,٢٥

$$= ٠,٥٠٠٠ - ٠,٣٩٤٤$$

$$= ٠,١٠٥٦$$

أى أن حوالى ١١٪ من الطلاب تزيد درجاتهم عن ٨٠ .

( ثالثاً )  $ل = (s \geq ٨٠) ل = (١,٢٥ \geq ع)$

= مساحة المنطقة التى على يسار العدد ١,٢٥

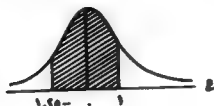
$$= ٠,٣٩٤٤ + ٠,٥٠٠٠ = ٠,٨٩٤٤$$



أى أن حوالى ٨٩٪ من الطلاب درجاتهم تساوى أو تقل عن ٨٠ .  
يلاحظ أن جواب أى من الاسئلة الثلاثة السابقة يمكن الحصول عليه من جوابى السؤالين الآخرين .

( رابعاً ) بوضع  $s = 40$  نجد أن  $E = \frac{60 - 40}{16} = 1,25$

بوضع  $s = 76$  نجد أن  $E = \frac{60 - 76}{16} = 1$



إذن ل  $(40 < s \leq 76)$  ل  $(1,25 < E \leq 1)$  = مساحة المنطقة فوق الفترة  $(1, 0)$  + مساحة المنطقة فوق الفترة  $(0, 1,25)$   
 $0,7307 = 0,3944 + 0,3413 = (0, 1,25 -)$

( خامساً )  $s = 70$  تعطى  $E = 0,9375 = 0,94$  تقريباً

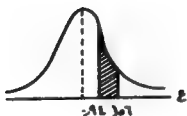
$s = 85$  تعطى  $E = 1,0625 = 1,06$

إذن ل  $(70 < s \leq 85)$  ل  $(1,06 < E \leq 0,94)$

= المساحة فوق الفترة  $(1,06, 0)$

- المساحة فوق الفترة  $(0,94, 0)$

$= 0,4406 - 0,3264 = 0,1142$



مثال ( ٤ - ٢ ) : مثال مشهور

للتغير المعتدل  $s$  الذى وسطه الحسابي  $\mu$  وانحرافه المعياري  $\sigma$  اثبت أن :

( أولاً ) احتمال وقوع قيم المتغير بين العددين  $\mu \pm \sigma$  هو ٠,٦٨٢٦

( ثانياً ) احتمال وقوع قيم المتغير بين العددين  $\mu \pm 2\sigma$  هو ٠,٩٥٤٤

( ثالثاً ) احتمال وقوع قيم المتغير بين العددين  $\mu \pm 3\sigma$  هو ٠,٩٩٧٤

الحل :

لكي نستخدم الجدول نحول المتغير  $x$  إلى المتغير  $z$  بواسطة التعويض :

$$z = (\mu - x) / \sigma$$

أولاً : بوضع :  $x = \mu - \sigma$  نجد أن  $z = \frac{\mu - (\mu - \sigma)}{\sigma} = 1$

وبوضع :  $x = \mu + \sigma$  نجد أن  $z = \frac{\mu - (\mu + \sigma)}{\sigma} = -1$

$\therefore P(\mu - \sigma < x < \mu + \sigma) = P(-1 < z < 1)$

$= 2 \times P(0 < z < 1)$  من الجدول

$= 2 \times 0.2420 = 0.4840$

ثانياً : بالمثل نجد أن :

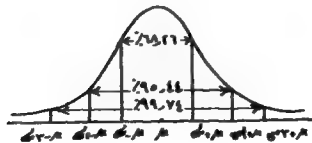
$P(\mu - 2\sigma < x < \mu + 2\sigma) = P(-2 < z < 2)$

$= 2 \times P(0 < z < 2) = 2 \times 0.4772 = 0.9544$

ثالثاً :  $P(\mu - 3\sigma < x < \mu + 3\sigma) = P(-3 < z < 3)$

$= 2 \times P(0 < z < 3) = 2 \times 0.4987 = 0.9974$

وتصور هذه النتائج بيانياً كما في الشكل (٤ - ٤) الآتي :



الشكل (٤ - ٤) المساحات أسفل المنحنى الطبيعي



### مثال (٤ - ٣) : مثال مشهور

بنفس الطريقة نثبت العلاقات الهامة الآتية التي سنحتاج إليها في كثير من التطبيقات الإحصائية ، انظر السؤال (١ - ح) من تمارين (٤) .

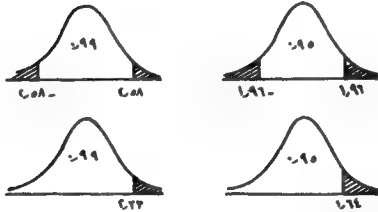
$$(١) \text{ ل } (١,٩٦ < \mathcal{E} \leq ١,٩٦) = ٠,٩٥$$

$$(٢) \text{ ل } (٢,٥٨ < \mathcal{E} \leq ٢,٥٨) = ٠,٩٩$$

$$(٣) \text{ ل } (١,٦٤ < \mathcal{E}) = ٠,٠٥$$

$$(٤) \text{ ل } (\mathcal{E} < ٢,٣٣) = ٠,٠١$$

وتمثل هذه الاحتمالات كما في الشكل (٤ - ٥) الآتي :



الشكل (٤ - ٥) بعض القيم الحرجة للمعيار المعتدل المعياري

### (٤ - ٣) الكشف عن الاعتدالية :

في كثير من الأحيان يبني التحليل الإحصائي لبيانات ناتجة من عينة على أساس افتراض أن المجتمع الذي سحبت منه العينة معتدل ، ولذلك ينبغي أن نتحقق من توفر هذا الافتراض قبل إجراء مثل هذا التحليل .

نفرض أن لدينا توزيعاً تكرارياً ذا فئات لعينة عشوائية وسطها الحسابي  $\bar{x}$  وانحرافها المعياري  $\sigma$  ونريد اختبار ما إذا كانت هذه العينة مأخوذة من مجتمع معتدل .

نتصور مجتمعاً معتدلاً له نفس الوسط الحسابي والانحراف المعياري للتوزيع التكراري المعلوم أى نأخذ  $\mu = \bar{x}$  ،  $\sigma = \sigma$  . إذا كانت العينة مأخوذة من هذا المجتمع فإن احتمال وقوع المتغير المعتدل في فئة مساوية لأى من الفئات التي ينقسم إليها التوزيع التكراري لا يجب أن يختلف كثيراً عن التكرار النسبي المشاهد في هذه الفئة . وعلى ذلك نقوم بحساب احتمالات وقوع المتغير المعتدل في جميع فئات التوزيع التكراري مستعينين في ذلك بمجدول المساحات . بضرب هذه الاحتمالات ( التكرارات النسبية ) في حجم العينة نحصل على ما يسمى بالتكرارات النظرية أو التوقعة المناظرة للتكرارات المشاهدة في العينة . نقارن بين التكرارات المشاهدة والتكرارات التوقعة لها فإذا كانت قريبة من بعضها بدرجة معقولة أى كانت هناك مطابقة حسنة بين التوزيع التكراري المشاهد والتوزيع التكراري النظري ، جاز لنا أن نعتبر أن المجتمع معتدل ، وإلا فهو ليس كذلك .

إن عملية إيجاد توزيع تكراري نظري بالطريقة المذكورة تسمى بعملية توفيق توزيع معتدل لتوزيع تكراري معلوم . وقد سبق أن مرت بنا فكرة التوفيق هذه في حالة كل من توزيع ذى الحدين وتوزيع بواسون . وكما سبق القول ، يعتمد الحكم الموضوعي على حسن المطابقة أو سوءها على أحد الاختبارات الإحصائية مثل اختبار  $\chi^2$  الذي سندرسه فيما بعد .

مثال ( ٤ - ٤ ) :

وفق التوزيع المعتدل للتوزيع التكراري الآتي ، وإذا كان هذا التوزيع لعينة عشوائية فاختر ما إذا كان من الممكن اعتبار أن المجتمع الذى سحبت منه هو مجتمع معتدل .

النفقة	٣١٥	٣٢٥	٣٣٥	٣٤٥	٣٥٥	٣٦٥	٣٧٥	٣٨٥	٣٩٥	٤٠٥
التكرار	٦	٦	١١	١٤	١٦	١٥	٨	١٠	٨	٦

الحل :

يتطلب توفيق التوزيع المعتدل أن نوجد الوسط الحسابي والانحراف المعياري للتوزيع التكراري المعطى لاستخدامهما في تقدير الوسط الحسابي والانحراف المعياري للمجتمع . كالاعتد نجد أن :

$$\text{الوسط الحسابي} = \bar{x} = \frac{1}{n} \sum x_i f_i = \frac{1}{100} (410 \times 6 + \dots + 330 \times 16 + 320 \times 6) = \frac{36470}{100}$$

$$= 364,7$$

$$\text{التباين} = s^2 = \frac{1}{n} \sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}$$

$$= \frac{1}{99} \left( \frac{36470^2}{100} - 410 \times 6^2 - \dots - 330 \times 16^2 - 320 \times 6^2 \right)$$

$$= 712,89$$

$$\text{الانحراف المعياري} = 26,7$$

نريد أن نختبر أن المجتمع الذي سحبت منه العينة هو مجتمع معتدل وسطه الحسابي ٣٦٤,٧ وانحرافه المعياري ٢٦,٧ . نحسب التكرارات المتوقعة كما في الجدول (١ - ٤) الآتي :

الجدول (٤ - ١)

توزيع معدل التكرار في الخلال (٤ - ٤)

التكرارات لمشاهدة لـ	التكرارات المتوقعة لـ $10 \times$	التكرارات النسبية للترتبة لـ	الفئات بقيم ع	الفئات بقيم س
٦	٦,٨١	٠,٠٦٨١	(١,٤٩) - ∞ -	٣٢٥ - ∞ -
٦	٦,٥٤	٠,٠٦٥٤	(١,١١) - (١,٤٩)	٣٣٥ - ٣٢٥
١١	٩,٦١	٠,٠٩٦١	(٠,٧٤) - (١,١١)	٣٤٥ - ٣٣٥
١٤	١٢,٩٨	٠,١٢٩٨	(٠,٣٦) - (٠,٧٤)	٣٥٥ - ٣٤٥
١٦	١٣,٦٦	٠,١٣٦٦	٠,٠١ - (٠,٣٦)	٣٦٥ - ٣٥٥
١٥	١٥,٥٧	٠,١٥٥٧	٠,٣٩ - ٠,٠١	٣٧٥ - ٣٦٥
٨	١٢,٤٧	٠,١٢٤٧	٠,٧٦ - ٠,٣٩	٣٨٥ - ٣٧٥
١٠	٩,٤٤	٠,٠٩٤٤	١,١٣ - ٠,٧٦	٣٩٥ - ٣٨٥
٨	٦,٣٧	٠,٠٦٣٧	١,٥١ - ١,١٣	٤٠٥ - ٣٩٥
٦	٦,٥٥	٠,٠٦٦٥	∞ - ١,٥١	∞ - ٤٠٥
١٠٠	١٠٠,٠٠	١,٠٠٠٠		

في العمود الأول من هذا الجدول نضع الفئات كما هي معطاة مع تعديل واحد وهو وضع - ∞ بدلا من الحد الأدنى للفتحة الأولى و + ∞ بدلا من الحد الأعلى

للفتة الأخيرة ، وذلك لأن التوزيع المعتدل هو توزيع متصل تقع قيمه بين  $-\infty$  ،  $+\infty$  وهذا التعديل من شأنه إدخال جميع هذه القيم دون أن يؤثر ذلك على التوزيع التكرارى المعطى .

وفي العمود الثاني نضع حدود الفئات بعد تحويلها إلى قيمها المعيارية بواسطة

$$\text{التعويض } ع = \frac{٣٦٤,٧ - ٣٢٥}{٢٦,٧} \text{ توظفة لاستخدام جدول المساحات أسفل المنحني}$$

المعتدل المعيارى ، فمثلا للفتة الأولى

$$٣ - = \infty \text{ تعطى } ع = - \infty .$$

$$١,٤٩ - = ٣ - = ٣٢٥ \text{ تعطى } ع = \frac{٣٦٤,٧ - ٣٢٥}{٢٦,٧} = ١,٤٩ -$$

وهكذا بالنسبة لبقية الفئات .

وفي العمود الثالث نضع التكرارات النسبية المتوقعة لـ التي تعبر عن احتمالات وقوع قيم المتغير ع في الفئات المناظرة ، وهذه الاحتمالات نوجدتها من جدول المساحات . فمثلا للفتتين الأولى والثانية :

$$ل (١,٤٩ \leq ع < \infty) = ٠,٥ - ٠,٤٣١٩ = ٠,٠٦٨١$$

$$ل (١,١١ \leq ع < ١,٤٩) = ٠,٤٣١٩ - ٠,٣٦٦٥ = ٠,٠٦٥٤$$

وهكذا بالنسبة لبقية الفئات .

ولما كانت هذه الاحتمالات هي بمثابة التكرارات النسبية في كل فئة فإننا للمقارنة بالتوزيع المعطى نضرب كلا منها في حجم هذا التوزيع وهو هنا ١٠٠ لنحصل على التكرارات المتوقعة قـ أى التكرارات التي نتوقعها في حالة كون المجتمع معتدلا . وهذه نضعها في العمود الرابع . أما العمود الخامس فيحمل التكرارات المشاهدة في العينة لتسهيل مقارنتها بالتكرارات المتوقعة .

ومن هذه المقارنة نشعر أن هناك مطابقة حسنة بين التوزيع التكرارى المشاهد والتوزيع التكرارى النظرى إذ لا تختلف التكرارات المشاهدة عن نظائرها المتوقعة في أغلب الفئات إلا قليلا مما يشير إلى أن المجتمع الذى أخذت منه العينة هو على الأرجح مجتمع معتدل ، وإن كان الحكم للموضوعى في ذلك يتطلب استخدام أحد الاختبارات الإحصائية المناسبة . انظر المثال (٦ - ٩) في البند (٦ - ٧ - ٢) .

#### (٤ - ٤) طريقة بيانية للكشف عن الاعتدالية :

هناك طريقة بيانية تستخدم كاختبار سريع للكشف عن دلالة المنحني التكرارى المشاهد ومدى انحرافه عن الاعتدالية ، ويشترط في هذه الطريقة أن تكون العينة عشوائية وكبيرة الحجم ( ن أكبر من ٥٠) . وتبنى فكرة هذه الطريقة على أن التوزيع المعتدل هو توزيع متماثل ذو تفرطح معين وبالتالي فإن أهم ما ينبغي التحقق منه في توزيع تكرارى لعينة هو مدى تماثله ومدى تفرطحه بالنسبة للتوزيع المعتدل .

وكل ما تتطلبه هذه الطريقة هو تكوين توزيع التكرارات المتجمعة المعوية للتوزيع التكرارى المعلوم ثم رسم النقط التي تمثل هذا التوزيع على ورق تقسيم الاحتمالات . فإذا وقعت هذه النقط على وجه التقريب على خط مستقيم يمكن توفيقه بالعين دل ذلك على أن المجتمع هو على الأرجح مجتمع معتدل ، وإلا فهو ليس كذلك . راجع الخاصة (هـ) من البند (٤ - ١) .

ولهذه الطريقة فائدة أخرى ، وهى أنه إذا ظهر لنا أن المجتمع معتدل أو قريب من الاعتدال فإننا نستطيع تقدير الوسط الحسابى والانحراف المعيارى لهذا المجتمع من المستقيم الذى وقفناه كالآتي :

(أ) الوسط الحسابى ( الوسيط في الواقع ) يقدر بالإحداثى السيني للنقطة التي على الخط المستقيم التي إحداثيها الصادى ٥٠ .

(ب) الانحراف المعيارى يقدر بنصف الفرق بين الإحداثيين السينيين للنقطتين اللتين إحداثييهما الصاديان ١٥,٩ ، ٨٤,١ .

مثال (٤ - ٥) :

استخدم الطريقة البيانية لاختبار اعتدالية المجتمع الذى سحبت منه العينة المذكورة في المثال (٤ - ٤) ، وإذا رأيت أن المجتمع معتدل فأوجد تقديراً لوسطه الحسابي وانحرافه المعياري .

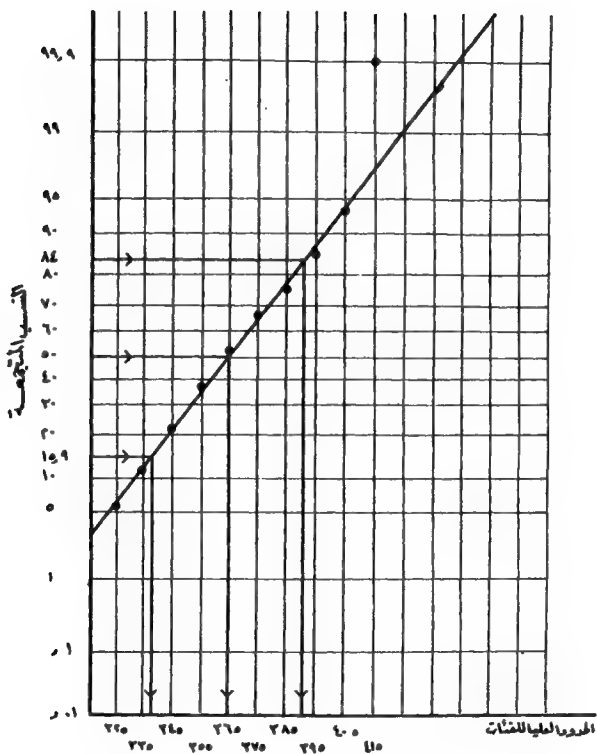
الحل :

نبدأ بإنشاء توزيع التكرارات المتجمعة المثوبة كما في الجدول (٤ - ٢) الآتي :

الجدول (٤ - ٢)

الحدود العليا للفئات	التكرار المتجمع	التكرار المتجمع %
$325 \geq$	٦	٦
$335 \geq$	١٢	١٢
$345 \geq$	٢٣	٢٣
$355 \geq$	٣٧	٣٧
$365 \geq$	٥٣	٥٣
$375 \geq$	٦٨	٦٨
$385 \geq$	٧٦	٧٦
$395 \geq$	٨٦	٨٦
$405 \geq$	٩٤	٩٤
$415 \geq$	١٠٠	١٠٠

نرسم النقط (٦ ، ٣٢٥) ، (١٢ ، ٣٣٥) ، ... ، (٩٤ ، ٤١٥) ، (١٠٠ ، ٤١٥) على ورق تقسيم الاحتمالات المعتدلة لنحصل على الشكل (٤ - ٦) الآتي :



الشكل (٤ - ٦) توزيع التكرارات النسبية للمنطقة نباتات اللؤلؤ (٤ - ٤)

موسمياً على ورق تقسيم الاحتمالات المعتدلة



بالتأمل في هذا الشكل نجد أن النقط تكاد تقع على خط مستقيم مما يشير إلى اعتدالية التوزيع . من الخط المرسوم نقدر الوسط الحسابي والانحراف المعياري للمجتمع كالآتي :

$$\begin{aligned}\text{الوسط الحسابي} &= 360 \\ \text{الانحراف المعياري} &= \frac{1}{4} (339 - 392) = 26,5\end{aligned}$$

#### (٤ - ٥) معالجة عدم اعتدالية التوزيع :

في التحليل الإحصائي للبيانات يتطلب الأمر في بعض الحالات أن يكون المجتمع معتدلاً ، فإذا لم يكن المجتمع معتدلاً نبحث عن تحويل مناسب يجعله معتدلاً أو قريباً من الاعتدال . ومن أكثر التحويلات استخداماً في هذا الصدد التحويل المسمى بالتحويل اللوغاريتمي logarithmic transformation الذى يحول المتغير  $x$  الذى لدينا إلى متغير  $y$  حيث  $y = \log x$  . ولا بأس من أخذ اللوغاريتمات العادية أى ذات الأساس ١٠ . وفي كثير من الحالات يفلح هذا التحويل في تعديل التوزيعات التكرارية المتوترة إلى البجين إلى توزيعات أكثر تماثلاً وبالتالي يكون قد عالج إلى حد ما عدم اعتدالية التوزيع . هذا مع ملاحظة أن عدم الاعتدالية لا تقترب عليه نتائج وخيمة إلا إذا كان التوزيع ملتوياً ملتوياً شديداً . على أنه إذا لم يوجد تحويل يصلح لتصحيح الاعتدالية فإننا نلجأ في تحليل البيانات إلى طرق لا تتطلب شرط الاعتدالية وهذه الطرق تسمى بالطرق غير البارامترية non-parametric methods وهى طرق لا تعتمد على فرض توزيعات محددة للمجموعات أو المتغيرات التى ندرسها . انظر الفصل الرابع عشر .

#### ملاحظة عن التحويلات :

(١) يستخدم التحويل اللوغاريتمي أيضاً في تحويل نموذج من النوع الضربى مثل  $y = x_1 x_2 x_3$  إلى نموذج من النوع الخطى  $\log y = \log x_1 + \log x_2 + \log x_3$

الذى هو أسهل تناولا ، وذلك بوضع  $ص = لو$  ،  $ص = س$  ،  $لو = س$  ... الخ ، وهذا ما نفعله أحيانا في موضوع تحليل الانحدار . كما يستخدم التحويل اللوغاريتمى حين يكون الوسط الحسابي  $\mu$  للمجتمع مرتبطاً ارتباطاً موجياً بالتباين  $\sigma^2$  فهو يحول المتغير الذى لدينا إلى متغير آخر يكون فيه هذان الدليلان مستقلين .

(٢) هناك تحويل آخر يسمى بتحويل الجذر التربيعى square-root transformation حيث نضع  $ص = \sqrt{س}$  . ويستخدم هذا التحويل للبيانات التى تنتج عن العد ويكون توزيعها بواسونيا حيث يكون الدليلان  $\mu$  ،  $\sigma^2$  غير مستقلين ( إذ نعلم أن  $\mu = \sigma^2$  ) ويفلح هذا التحويل في جعلهما مستقلين . وإذا احتوت البيانات على أصفار يفضل استخدام التحويل  $ص = \sqrt{س + \frac{1}{4}}$  .

(٣) من التحويلات الشهيرة أيضاً التحويل الزاوى angular transformation حيث نضع  $ص = \arcsin \sqrt{س}$  ويستخدم حين تكون البيانات مؤلفة من نسب مئوية . وفي توزيع ذى الحدين الذى دليلاه  $ن$  ،  $ح$  نعلم أن الوسط الحسابي  $\mu = ن ح$  والتباين  $\sigma^2 = ن ح (١ - ح)$  وبالتالي فإن التباين يكون دالة في الوسط الحسابي . إن التحويل الزاوى يوقف هذه الدالية . إلا أنه حين تكون النسب واقعة بين ٠,٣٠ ، ٠,٧٠ ، فإن هذا التحويل لا يكون له ضرورة .

(٤) إن التحويلات الثلاثة السابقة تغير العلاقة الدالية بين المتغيرات أى تغير النموذج الرياضى إلى نموذج آخر . غير أن هناك تحويلات لا تفعل هذا وإنما تستهدف تقنين المتغيرات عن طريق :

(أ) استبعاد وحدات القياس وذلك بالتحويل إلى مقياس نسبى لا يعتمد على وحدات القياس ،

(ب) جعل القيم الناتجة عن التحويل في المجموعات المختلفة تتساوى في أوساطها الحسابية وفي تبايناتها .

وأشهر هذه التحويلات يأخذ الصورة الآتية المسماه بالصورة المعيارية :

$$\frac{\bar{s} - s}{\bar{e}} = \bar{v}$$

حيث  $\bar{s}$  متوسط قيم  $s$  ،  $\bar{e}$  انحرافها المعياري .

ونظراً لأن هذا المقياس نسبي فإن القيم الصادية الناتجة تكون خالية من أى وحدة قياس ، كما أن هذا التحويل إذا أُجرى على قيم المتغير فى أى مجموعة فإنه يحول هذه القيم إلى قيم متوسطها يساوى صفراً وتباينها يساوى الواحد الصحيح .

#### تقارين (٤)

(١) للتوزيع المعتدل المعيارى وباستخدام جدول المساحات

(أ) أوجد كلا من الاحتمالات الآتية :

$$\begin{aligned} & \text{ل } (0 < e \leq 1,0) , \text{ ل } (1,65 \leq e) , \text{ ل } (e > 0,74) \\ & \text{ل } (1,25 \leq e < 2,00) . \end{aligned}$$

$$\begin{aligned} & \text{(ب) أوجد قيمتي أ ، ب بحيث ل } (e > \text{ب}) = 0,90 , \\ & \text{ل } (e < \text{أ}) = 0,48 \end{aligned}$$

$$\begin{aligned} & \text{(ج) أثبت أن ل } (1,96 \leq e < 1,96) = 0,90 , \text{ ل } (e \leq 1,64) \\ & = 0,05 \end{aligned}$$

$$\text{ل } (2,08 \leq e < 2,08) = 0,99 , \text{ ل } (e \leq 2,33) = 0,01 ,$$

(٢) للتوزيع المعتدل الذى وسطه الحسابى ٥٠ وانحرافه المعيارى ٥ أوجد كلا من

$$\text{ل } (30,5 < s \leq 60,5) , \text{ ل } (50 \leq s < 40) .$$

(٣) فى مجتمع معين المعروف أن نسب الذكاء تتوزع توزيعاً معتدلاً وسطه الحسابى ١٠٤,٦ وانحرافه المعيارى ٣,١٥

(أ) أوجد نسبة الأفراد الذين تقع نسب ذكائهم بين ٩٠ ، ١٢٠

(ب) أوجد نسبة الأفراد الذين تزيد نسب ذكائهم عن ١١٠

(٤) التوزيع الآتي هو التوزيع التكرارى لعينة عشوائية مأخوذة من مجتمع ما .

٢٧,٥	٣٢,٥	٣٧,٥	٤٢,٥	٤٧,٥	٥٢,٥	٥٧,٥	٦٢,٥
٥	٦	١٣	٨	١٦	١٤	٧	١

(أ) أثبت أن الوسط الحسابي يساوى ٤٧,٠٧١ وأن الانحراف المعياري يساوى ٨,٩٤٨ .

(ب) وفق توزيعاً معتدلاً لهذا التوزيع واذكر رأيك فيما إذا كان بالإمكان اعتبار أن المجتمع معتدل .

(ج) استخدم الطريقة البيانية لاختبار اعتدالية المجتمع .

#### (٤ - ٦) تقريب توزيع ذى الحدين بتوزيع معتدل :

في البند (٣ - ٤ - ١) رأينا أنه إذا كان  $n$  متغيراً عشوائياً ذا توزيع ذى حدين دليلاً  $n$  ، ح يجوز تقريب هذا التوزيع بتوزيع بواسون متوسطه يساوى متوسط توزيع ذى الحدين بشرط أن يكون حجم العينة  $n$  كبيراً وأن يكون الاحتمال  $p$  صغيراً حيث يكون التوزيع ملتوياً إلى اليمين . يجوز تحت شروط أخرى تقريب توزيع ذى الحدين : حد (  $n$  ،  $p$  ) بتوزيع معتدل متوسطه  $\mu = np$  حيث  $n$  أى يساوى متوسط توزيع ذى الحدين ، وتباينه  $\sigma^2 = np(1-p)$  حيث  $n$  أى يساوى تباين ذى الحدين بشرط أن تكون  $n$  كبيرة وأن تكون  $p$  قريبة من العدد  $\frac{1}{2}$  حيث يكون التوزيع متاثلاً بالتقريب . تحت هذين الشرطين يقترب توزيع الإحصاءة

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \quad \text{حيث } k = 1 - \alpha$$

من التوزيع المعتدل المعياري : مع (١٤٠) كلما زاد العدد  $n$  .

ومن الناحية العملية وجد أن هذا التقريب يكون جيداً أى يمكن التجاوز عن الخطأ الناشئ عنه إذا توفر أحد الشرطين الآتين :

(أ) إذا كانت كل من  $n$  و  $h$  أكبر من خمسة

أو (ب) إذا كانت  $n \leq 10$  ومعامل الالتواء أصغر من ٠,٢ .

ونظراً لأن توزيع ذى الحدين هو توزيع لمتغير وثاب بينما التوزيع المعتدل هو توزيع لمتغير متصل فإننا لعلاج ذلك فى عملية التقريب يجب أن نعتبر أن كل قيمة  $s$  من قيم المتغير ذى الحدين ممتدة نصف وحدة من اليسار ونصف وحدة من اليمين فمثلاً العدد  $s = 4$  نعتبر أنه الفترة (٣,٥ ، ٤,٥) ، والعدد  $s = 4,٧$  نعتبر أنه الفترة (٤,٦٥ ، ٤,٧٥) ، وإذا أردنا مثلاً إيجاد الإحتمال لـ  $(10 \leq s \leq 11)$  نعتبر أن  $s$  يتوزع ذى حدين فإننا نحسب الاحتمال التقريبي لـ  $(10,5 \leq s \leq 11,5)$  باستخدام جداول التوزيع المعتدل .

مثال (٤ - ٦)

ألقيت حجرة نرد منتظمة عشوائياً ١٠ مرات . أوجد احتمال الحصول على الصورة في ٣ أو ٤ أو ٥ أو ٦ مرات .

الحل :

نظراً لأن حجرة النرد منتظمة والرمية عشوائية فإن توزيع عدد مرات ظهور الصورة هو توزيع ذو حدين : حد (١٠ ، ٠,٥) ودالة كتلة الاحتمال تكون كالآتي :

$$d(s) = \binom{10}{s} (0,5)^s (0,5)^{10-s} = \binom{10}{s} (0,5)^{10}$$

حيث  $s = 0, 1, 2, \dots, 10$

$$\therefore d(s) = d(3) + d(4) + d(5) + d(6)$$

$$= (0,5) \cdot ( \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} )$$

$$= 0,7734$$

الحل التقريبي :

بما أن الشرط (ب) متوفر إذ أن  $0 \leq 10$  ومعامل الالتواء = صفر  $> 2,0$  ،  
 ( مع ملاحظة أن التوزيع متماثل تماماً لأن  $\frac{1}{4} = \frac{1}{4}$  ) يمكن التقريب بتوزيع معتدل  
 متوسطه  $\mu = 5$  ،  $\sigma = 0,5 \times 10 = 5$  وتباينه  $\sigma^2 = 0,5 \times 10 = 5$  ، ويكون للمتغير  
 $2,0$  واذن انحرافه المعياري يساوي  $\sqrt{2,0} = 1,41$  ويكون للمتغير

$$z = \frac{5 - 2,0}{1,41}$$

توزيع معتدل معياري على وجه التقريب . وهنا نعتبر أن العدد 3 هو الفترة  $(2,0)$  ،  
 $(3,0)$  ، وأن العدد 4 هو الفترة  $(3,0)$  ، وهكذا ... ويكون المطلوب إيجاد  
 احتمال  $L(2,0 < S < 6,0)$  في التوزيع المعتدل مع  $(1,41, 0)$

$$\text{بوضع } S = 2,0 \text{ نجد أن } z = \frac{5 - 2,0}{1,41} = 2,13$$

$$\text{بوضع } S = 6,0 \text{ نجد أن } z = \frac{5 - 6,0}{1,41} = -0,71$$

$$L(2,0 < S < 6,0) = L(0,71 < z < 2,13) = 0,7718$$

ومن الواضح أن هذه القيمة قريبة جداً من القيمة المضبوطة 0,7734.

## (ثانيا ) التوزيع المعتدل اللوغاريتمى

إن التوزيع المعتدل اللوغاريتمى هو مثال آخر نماذج الاحتمال المتصلة ، وقد سمى كذلك لأن التحويل  $ص = لو س$  يحول المتغير الذى يصفه هذا التوزيع إلى متغير ذى توزيع معتدل . ومن الظواهر التى يصلح لها هذا النموذج بعض الظواهر الجيولوجية كتلك المتعلقة بأوزان وأعداد بعض أنواع الصخور الرسوبية ، وبعض الظواهر الاقتصادية كدخول الأفراد وخاصة الدخول ذوات القيم الصغيرة .

ويعرف هذا التوزيع بواسطة دالة كثافة الاحتمال الآتية :

$$د(ص) = \frac{1}{\sqrt{\theta} \sqrt{1 - \theta}} \quad \text{أو} \quad \frac{1}{\sqrt{\theta} \sqrt{1 - \theta}} \quad (1) \quad 0 < \theta$$

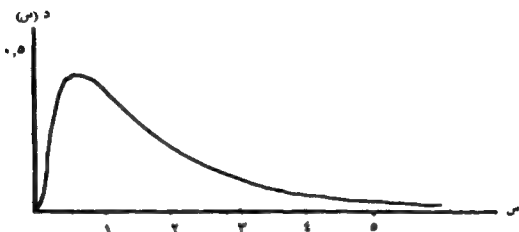
حيث اللوغاريتم للأساس  $هـ$  وحيث  $\theta$  ، بارامتران إذا علمت قيمتهما تحدد التوزيع تحديدا تاما .

### (٤ - ٦) بعض خصائص التوزيع

(١) المنحنى الممثل للدالة (٤) هو منحنى ذو قمة واحدة وملتوى إلى اليمين ، ويختلف شكله باختلاف قيمتى البارامترين  $\theta$  ،  $\gamma$  ، فمثلا يأخذ الشكل المبين بالشكل (٤-٧) حين تأخذ  $\gamma$  القيمة صفر وتأخذ  $\theta$  القيمة ١ .

(ب) التحويل  $ص = لو س$  حيث اللوغاريتم للأساس  $هـ$  يحول المتغير  $ص$  إلى متغير  $ص$  له توزيع معتدل وسطه الحسابى  $\gamma$  وانحرافه المعيارى  $\theta$  ، وبالتالي يكون المتغير

$$\frac{\gamma - \gamma}{\theta} = \epsilon$$



الشكل (٧-٤)

المنحنى المعتدل القوطي:  $\gamma = 0$  و  $\theta = 1$

هو متغير معتدل معياري ( وسطه الحسابي صفر وتباينه ١ ) . وبناء على ذلك يمكن استخدام جدول المساحات أسفل المنحنى المعتدل للمياري ، وهو الجدول (٦) بملحق هذا الكتاب ، في إيجاد الاحتمالات والقيم الحرجة للمتغير  $س$  كما في المثالين الآتيين :

مثال (٦-٤)

إذا كان  $س$  متغيراً معتدلاً لوغاريتمياً ذليلاً  $\gamma = 1$  ،  $\theta = 0$  فأوجد الاحتمال  $L(س > ٣٥,٥)$  .

الحل :

بوضع  $ص = لو(س)$  يكون للمتغير  $ع = \frac{١ - س}{٢}$  توزيع معتدل معياري.

$$L(س > ٣٥,٥) = L(لو س > لو ٣٥,٥) = L\left(\frac{١ - س}{٢} > \frac{١ - ٣٥,٥}{٢}\right)$$

$$L(ع > ١,٢٨٤٨) = L\left(\frac{١ - ٣,٥٦٩٥}{٢} > ع\right)$$

$$= ٠,٨٩٩٧ \text{ (من جدول المساحات أسفل المنحنى المعتدل)}$$



#### مثال (٧-٤)

إذا كان  $\theta$  متغيراً معتدلاً لوغاريتمياً دليلاً  $\gamma = 2$  ،  $\theta = 0,5$  ، فأوجد قيمة  $\alpha$  بحيث  $L(\alpha > 1) = 0,90$ .

الحل :

$$L(\alpha > 1) = L(\text{لو } \alpha > \text{لو } 1) = L\left(\frac{\text{لو } \alpha - 1}{0,5} > \frac{\text{لو } 1 - 1}{0,5}\right)$$

$$L(\alpha > 1) = L(\alpha > 1) = 0,90$$

$$\text{من جدول المساحات نجد أن } \alpha = 1,28 = \frac{\text{لو } \alpha - 1}{0,5}$$

$$\therefore \text{لو } \alpha = 1,28 \therefore \alpha = 1,43$$



## الفصل الخامس

### توزيعات خاصة

#### SPECIAL DISTRIBUTIONS

كل من التوزيعات الثلاثة الآتية هو توزيع احتمال لمتغير عشوائي متصل مركب بطريقة معينة من عدد من المتغيرات العشوائية المعتدلة . وهذه التوزيعات لا تستخدم كنماذج احتمال للمجتمعات كما هو الحال في التوزيعات التي عرضت في الفصولين السابقين ، وإنما تبرز من خلال التحليل الإحصائي للعينات وتبني عليها اختبارات إحصائية ذات أهمية قصوى في عملية الاستدلال الإحصائي كما سنرى في الفصل التالي . ومن الناحية التطبيقية يهتما بصفة خاصة في دراسة هذه التوزيعات أمرين هما :

(١) الشكل الهندسي العام لكل توزيع .

(٢) كيفية استخدام الجداول لإيجاد الاحتمالات والقيم الحرجة .

#### STUDENT t- DISTRIBUTION

(١-٥) توزيع ت :

يقال لمتغير عشوائي  $t$  إن له توزيع  $t$  إذا وقط إذا كانت دالة كثافة احتمال معرفته بالقاعدة :

$$(1) \quad d(s) = \frac{(1+s)^{\frac{1}{\nu}}}{\left(\frac{\tau}{s} + 1\right)^{\frac{1}{\nu}}} \frac{1}{\nu! (2-s)^{\frac{1}{\nu}} \cdot \sqrt{\nu}} = (s)$$

حيث  $0 < s < \infty$  ،  $0 < \nu < \infty$  ،  $1 = \nu$  ،  $2 = \nu$  ،  $3 = \nu$  ...

وهذا التوزيع له دليل واحد هو  $\nu$  يسمى بعدد درجات الحرية للتوزيع ، ولهذا فالتوزيع يتحدد تماماً إذا علمت قيمة  $\nu$  .



الشكل (١-٥) منحنى توزيع للغير ت

والمنحنى الممثل للدالة  $s = d(s)$  المعرفة في (١) هو منحنى ذو قمة واحدة ومتماثل حول المستقيم  $s = \text{صفر}$  .

(٢) الوسط الحسابي للتوزيع :  $\mu = \text{صفر}$

(٣) ، تباین التوزيع :  $\sigma^2 = \frac{\nu}{2-s}$  حيث  $\nu < 2$

وجدير بالذكر أن توزيع ت يقترب من التوزيع المعتدل المعيارى كلما اقتربت درجات الحرية  $\nu$  من اللانهاية .

جدول القيم الحرجة :

الجدول (٧) بملحق هذا الكتاب يعطى القيمة الحرجة  $\text{critical value}$  للمتغير  $t$  وهى تلك القيمة الموجبة التي تجعل مساحة المنطقة أسفل المنحنى وخارج الفترة

(ت. ، ت.) مساوية لقيمة معينة  $\alpha$  عند درجة الحرية  $n$  . أي أن العدد  $\alpha$  يعبر عن المساحة عند ذيل المنحني ( $\alpha$  عند كل ذيل) وبالتالي فهو يعبر عن احتمال وقوع قيم المتغير ت خارج الفترة المذكورة . (انظر الشكل ١-٥) . ونكتب هذا رمزياً كالآتي :

$$\alpha = ( | ت | < ت ) \quad (٤)$$

في هذا الجدول كتبت درجات الحرية في كل من العمودين الهامشين الواقعين في يمين ويسار الجدول ، وكتبت الاحتمالات  $\alpha = ٠,٠٠١ , ٠,٠١ , ٠,٠٢ , ٠,٠٥ , ٠,١ , ٠,٢ , ٠,٤ , ٠,٥ , ٠,٩$  في الهامش الأتقى الذى بأعلى الجدول . أما القيم الحرجة فمعدونة في الخلايا التي بقلب الجدول . وحين تكون درجة الحرية  $n$  معلومة لنا ، نستطيع استخراج مايلى :

- (١) القيمة الحرجة ت : إذا أعطيت قيمة الاحتمال  $\alpha$  . ونجد تلك القيمة عند نقطة التقاء الصف الذى به درجة الحرية  $n$  والعمود الذى به قيمة  $\alpha$  .
- (٢) قيمة الاحتمال  $\alpha$  إذا أعطيت القيمة الحرجة . ولإيجاد تلك القيمة نبحث في الصف الذى به درجة الحرية عن القيمة الحرجة المعطاة فتكون  $\alpha$  هي العدد الذى يعلو العمود الذى به هذه القيمة .

مثال (١-٥) :

ليكن  $m$  متغيراً له توزيع ت بدرجات حرية عددها ٨ . من الجدول نجد مايلى :

- (١) إذا كانت  $\alpha = ٠,٠٥ = ( | ت | < ت )$  فإن القيمة الحرجة ت = ٢,٣٠٦
- (ب) الاحتمال  $\alpha = ٠,٠٢ = ( | ت | < ٢,٨٩٦ )$
- (ج) الاحتمال  $\alpha = ٠,٠١ = ( ت < ٢,٨٩٦ )$  مساحة الذيل الأيمن فقط .

تقليد :

(٥)

نكتب  $T_{100}$

للتعبير عن القيمة الحرجة  $T$  في توزيع  $T$  عند درجة الحرية  $n$  بحيث يكون مجموع مساحتي المنطقتين عند الذيلين يساوي  $\alpha$  ، فمثلا :

$$T_{100,0.05} = 1,96 \quad , \quad T_{100,0.01} = 2,365 \quad , \quad T_{100,0.001} = 3,090$$

THE CHI-SQUARE DISTRIBUTION

(٢-٥) توزيع  $\chi^2$  :

يقال لتغير عشوائي  $\chi^2$  إن له توزيع  $\chi^2$  (تنطق كاي تربيع) إذا وقطع إذا كانت دالة كثافة احتماله معرفة بالقاعدة :

$$f(x) = \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad (٦)$$

حيث  $0 < x < \infty$  ،  $0 < n < \infty$  ،  $n = 1, 2, 3, \dots$

وهذا التوزيع له دليل واحد هو  $n$  يسمى بعدد درجات الحرية للتوزيع ، ولهذا فالتوزيع يتحدد تماماً إذا علمت قيمة  $n$  .

والمنحنى الممثل للدالة  $f(x) = f(n)$  للمعرفة في (٦) وعندما  $n < 2$  يكون منحنى ذو قمة واحدة وملتوى إلى اليمين .



الشكل (٢-٥) منحنى توزيع الطور  $\chi^2$  ،  $(n < 2)$

- (٧) الوسط الحسابي للتوزيع :  $\mu = \nu$
- (٨) ، تباين التوزيع :  $\sigma^2 = \nu$

جدول القيم الحرجة :

الجدول (٨) يملحق هذا الكتاب يعطى القيمة الحرجة  $\chi^2$  للمتغير  $\chi^2$  وهي القيمة التي تجعل مساحة المنطقة أسفل المنحني وفوق الفترة  $(\chi^2, \infty)$  مساوية لقيمة معينة  $\alpha$  عند درجة الحرية  $\nu$ . أى أن العدد  $\alpha$  يعبر عن المساحة عند الذيل المتتوى، وبالتالي فهو يعبر عن احتمال وقوع قيم المتغير  $\chi^2$  على يمين العدد  $\chi^2$  (انظر الشكل ٢-٥). ونكتب هذا رمزياً كالآتي :

$$(٩) \quad \alpha = P(\chi^2 < \chi^2)$$

في هذا الجدول كتبت درجات الحرية والاحتمالات  $\alpha$  والقيم الحرجة بنفس الطريقة التي كتبت بها في جدول ت، غير أن  $\alpha$  تأخذ القيم ٠,٠٠٥ ، ٠,٠١ ، ٠,٠٥ ، ٠,١ ، ٠,٢٥ ، ٠,٥ ، ٠,٩ ، ٠,٩٧٥ ، ٠,٩٩٥ .  
وحيث تكون درجة الحرية  $\nu$  معلومة لنا، نستطيع من الجدول استخراج القيمة الحرجة  $\chi^2$  إذا أعطيت قيمة الاحتمال  $\alpha$  أو استخراج الاحتمال  $\alpha$  بمعلومة القيمة الحرجة  $\chi^2$ .

مثال (٢-٥) :

ليكن  $\nu$  متغيراً له توزيع  $\chi^2$  بـ درجات حرية عددها ١٠. من الجدول نجد مايلي :

$$(أ) \quad \text{إذا كانت } \nu = 10 \text{ فإن القيمة الحرجة } \chi^2 = 15,987$$

$$(ب) \quad \text{الاحتمال } P(\chi^2 < 18,31) = 0,05$$

$$(ج) \quad \text{الاحتمال } P(4,87 < \chi^2 < 18,31) = 0,90 - 0,05 = 0,85$$

تقليد :

(١٠)

نكتب  $\chi^2_{\alpha}$

للتعبير عن القيمة الحرجة  $\chi^2$  في توزيع  $\chi^2$  عند درجة الحرية  $\nu$  بحيث تكون مساحة المنطقة التي إلى يمينها مساوية للعدد  $\alpha$  :

$$\chi^2_{0.05, 10} = 15.99, \chi^2_{0.01, 10} = 21.01, \chi^2_{0.001, 10} = 23.58$$

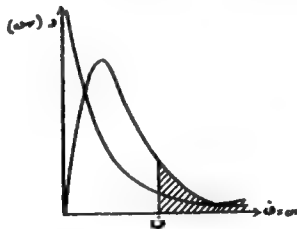
### THE F-DISTRIBUTION

(٣-٥) توزيع ف :

يقال لمتغير عشوائي  $F$  إن له توزيع ف إذا وقط إذا كانت دالة كثافة احتماله معرفة بالقاعدة :


$$(11) \quad \frac{1 - \frac{1}{1 + \frac{\nu_2}{\nu_1} x}}{\left(1 + \frac{\nu_2}{\nu_1} x\right)^{\frac{\nu_1}{2} + 1}} \times \frac{\frac{\nu_2}{2} \left(\frac{\nu_2}{2} - 1\right) \dots \left(\frac{\nu_2}{2} - \nu_1 + 1\right) \frac{1}{\nu_1}}{\frac{\nu_2}{2} \left(\frac{\nu_2}{2} - 1\right) \dots \left(\frac{\nu_2}{2} - \nu_2 + 1\right) \frac{1}{\nu_2}} = f(x)$$

حيث  $0 < x < \infty$  ،  $\nu_1$  ،  $\nu_2$  عدنان صحيحان موجبان .  
وهذا التوزيع له دليلان هما  $\nu_1$  ،  $\nu_2$  يسميان بعددي درجات الحرية ، ويتحدد التوزيع تماماً إذا علمت قيمتي هذين الدليلين .



الشكل (٣-٥) منحنى توزيع الف



والمنحني الممثل للدالة  $\chi^2 = D$  (س) المعرفة في (١١) يختلف شكله بحسب قيمتي ٢ ،  $\nu$  فهو يأخذ الشكل  إذا كانت ٢ ،  $\nu$  صغيرتين جداً إلا أنه يصبح محدباً وملتوياً إلتواء شديداً إلى اليمين كلما زادت قيمتا ٢ ،  $\nu$  .

$$\text{الوسط الحسابي للتوزيع} = \mu = \frac{\nu}{2-\nu} \quad \text{حيث } \nu < 2 \quad (١٢)$$

### جدول القيم الحرجة :

الجدول (٩) يملحق هذا الكتاب يعطى القيمة الحرجة ف. للمتغير ف وهي تلك القيمة التي تجعل مساحة المنطقة أسفل المنحني وفوق الفترة (ف. ،  $\infty$ ) مساوية لقيمة معينة  $\alpha$  عند درجتي الحرية ٢ ،  $\nu$  . أى أن العدد  $\alpha$  يعبر عن المساحة عند الذيل الملتوى وبالتالي فهو يعبر عن احتمال وقوع قيم ف على يمين العدد ف . (انظر الشكل ٥-٣) ونكتب هذا رمزياً كالآتي :

$$\text{ل(ف < .ف)} = \alpha \quad (١٣)$$

ويختلف تركيب هذا الجدول عن جدول  $\chi^2$  إذ يحمل الهامش الألقى العلوى درجات الحرية ٢ ويحمل كل من العمودين الهامشين على جانبي الجدول درجات الحرية  $\nu$  وعند كل من هذه الدرجات وضعت ثلاث قيم للعدد  $\alpha$  هي ٠,٠٥ ، ٠,٠٢٥ ، ٠,٠١ وعند تقاطع كل عمود ٢ مع كل صف  $\nu$  توجد ٣ قيم حرجة تناظر تلك القيم الثلاث .

وحين تكون درجتا الحرية ٢ ،  $\nu$  معلومتين نستطيع من الجدول استخراج القيمة الحرجة ف. بمعلومية الاحتمال  $\alpha$  أو استخراج قيمة الاحتمال  $\alpha$  بمعلومية القيمة الحرجة .

مثال (٥-٣) :

ليكن  $\alpha$  متغيراً له توزيع ف بلرجي حرية ٧ ، ٩ .

(أ) إذا كانت ل (ف < ف) = ٠,٠٥ فإن ف = ٣,٢٩

(ب) إذا كانت ل (ف < ف) = ٠,٠١ فإن ف = ٥,٦١

(ج) الاحتمال ل (ف < ٤,٢) = ٠,٠٢٥

ملاحظة :

لا يعطى الجدول القيم الحرجة إلا عند ثلاث قيم للاحتمال  $\alpha$  هي ٠,٠٥ ، ٠,٠١ ، ٠,٠٢٥ ولكن يمكننا أيضاً إيجاد تلك القيم عند الاحتمالات المكملة ٠,٩٩ ، ٠,٩٧٥ ، ٠,٩٥ باستخدام النظرية الآتية :

إذا كان  $\alpha$  متغيراً عشوائياً له توزيع ف بلرجي حرية ٢ ، ن فإن المتغير  $\frac{1}{\alpha}$  يكون له توزيع ف بلرجي حرية ٢ ، ١ . ويمكن أن نكتب هذه النظرية كالآتي :

$$\frac{1}{\alpha} = F_{(1-\alpha)(n-1)}^n$$

مثال (٥-٤) :

ليكن  $\alpha$  متغيراً له توزيع ف بلرجي حرية ٥ ، ٩ أوجد قيمة ف بحيث ل(ف < ف) = ٠,٩٥

الحل :

نلاحظ أن الاحتمال ٠,٩٥ ليس له وجود بالجدول أما الاحتمال المكمل ٠,٠٥ فموجود به . التباينة ف < ف تكافئ التباينة  $\frac{1}{\alpha} > \frac{1}{\alpha}$  مع ملاحظة أن ف ، ف موجبان .

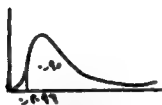
إذن لـ (ف < ف) = ل =  $\left(\frac{1}{3} > \frac{1}{3}\right) = 0,90$  فرضاً .

إذن لـ  $\left(\frac{1}{3} \leq \frac{1}{3}\right) = 0,00$

من النظرية يكون لدينا متغير له توزيع ف بدرجتي حرية ٩ ، ٥

من الجدول نجد أن  $\frac{1}{3} = 4,77$

إذن ف =  $1 = 4,77 \div 0,2096$



الشكل (٥-٤)

تقليد :

(١٤)

نكتب

ف [٥,٢]

للتعبير عن القيمة الحرجة في توزيع ف عند درجتي الحرية م ، ن بحيث تكون مساحة المنطقة على يمينها مساوية للعدد  $\alpha$  ، فمثلاً :

ف [٨,٥] = ٣,٦٩ ، ف [٥,٨] = ٤,٨٢ ، ف [٤,١٥] = ١٤,٢

لاحظ أن العدد الأول م يحدد العمود والعدد الثاني ن يحدد الصف .

## تمارين (٥)

(١) أوجد كلا من القيم الحرجة الآتية :

( أ ) ت [١٧] ، ٠,٠٥ ، ت [٤٠] ، ٠,٠٢ ، ت [١٦] ، ٠,٠١ ، ت [٥٣] ، ٠,٠٥

( ب )  $\chi^2_{[٢٠], ٠,٠٥}$  ،  $\chi^2_{[٣٥], ٠,٠١}$  ،  $\chi^2_{[٣٧], ٠,٠١}$  ،  $\chi^2_{[١١], ٠,٠٥}$

( ح ) ف [٨٦] ، ٠,٠٥ ، ف [٦٥] ، ٠,٠٥ ، ف [١٠٠] ، ٠,٠١ ، ف [١٤٣] ، ٠,٠٥

(٢) أوجد كلا من الاحتمالات الآتية :

$$(١) \text{ ل } | \text{ ت } | < (٢, ٢٠١)$$

$$\text{، ل } ( \text{ ت } < (٢, ٢٠١)$$

$$(ب) \text{ ل } ( \text{ خ } < (٥٢, ١٩)$$

$$\text{، ل } ( \text{ خ } < (١٩, ٣٤)$$

$$(ج) \text{ ل } ( \text{ ف } < (١٠, ٧)$$

$$\text{، ل } ( \text{ ف } < (٨, ٥)$$

$$\text{حيث } \text{ ن } = ١١$$

$$\text{حيث } \text{ ن } = ١١$$

$$\text{حيث } \text{ ن } = ٣١$$

$$\text{حيث } \text{ ن } = ٢٠$$

$$\text{حيث } \text{ ن } = ٥ ، \text{ ن } = ٦$$

$$\text{حيث } \text{ ن } = ٤ ، \text{ ن } = ٣٠$$

## الفصل السادس

### نظرية العينات

#### THEORY OF SAMPLING

##### (٦-١) نظرية العينات :

تبحث نظرية العينات في العلاقات بين المجتمعات والعينات المأخوذة من هذه المجتمعات ، وقد حوت من النظريات والتوزيعات والأساليب ما يمكننا من تقدير أدلة المجتمعات أو مقارنتها أو إصدار قرارات أو تنبؤات بشأنها عن طريق دراسة وتحليل عينات مأخوذة منها مما يدخل تحت موضوع الاستدلال الإحصائي .

##### . STATISTICAL INFERENCE

ويقصد بالاستدلال الإحصائي أى إجراء يستخدم نظرية الاحتمال في إصدار قرارات عن مجتمع أو عدة مجتمعات عن طريق عينات مأخوذة منها مع تحديد درجة الثقة في هذه القرارات . ولعملية الاستدلال الإحصائي مجالان رئيسيان هما مجال تقدير بارامترات المجتمعات ومجال اختبار الفروض الإحصائية ، على أنه من الشروط الأساسية في هذه العملية أن تكون العينات عشوائية لأن جميع نظريات الاحتمال التي تعتمد عليها مؤسسة على فرض العشوائية . كما أنه كلما كبر حجم العينة كلما كان الاستدلال أكثر دقة .

ومن بين المسائل التي تبرز في الأبحاث التطبيقية وتحتاج إلى عملية الاستدلال الإحصائي ما يلي :

- (أ) اختبار صواب أو خطأ فروض مطروحة عن أدلة المجتمعات .
- (ب) تقديم متوسطات وتباينات المجتمعات وغيرها من الأدلة .
- (ج) الكشف عما إذا كانت الفروق المشاهدة في العينات هي فروق راجعة إلى الصدفة أو تقلبات العينات أو هي فروق جوهرية تدل على وجود اختلاف حقيقي بين المجتمعات التي أخذت منها هذه العينات .
- (د) الكشف عن تأثير واحد أو أكثر من العوامل أو المعالجات على متغير ما أو ظاهرة معينة .
- (هـ) تقدير درجة ونوع العلاقة أو الارتباط بين المتغيرات وإصدار تنبؤات عنها .
- وسوف نتدارس هذه المسائل وغيرها في هذا الفصل وما يليه من فصول ، بعد تقديم بضعة مفاهيم ونظريات تتطلبها الدراسة الواعية لتلك المسائل .

## (٦-٢) توزيعات المعاينة :

في نظرية العينات نميز بين ثلاثة أنواع من التوزيعات .

Population Distribution	(١) توزيع المجتمع
Sample Distribution	(٢) توزيع العينة
Sampling Distribution	(٣) توزيع المعاينة

ولبيان الفرق بين هذه التوزيعات نفرض أننا نرغب في إيجاد الوسط الحسابي لدخل العائلة في مجتمع مدينة مؤلفة من ٢٠٠٠ عائلة . إذا أمكننا معرفة دخول جميع عائلات المدينة ووضعنا هذه الدخول في توزيع تكرارى فإن هذا التوزيع يسمى بتوزيع المجتمع للدخول . ونستطيع بالطبع أن نحصل مباشرة على الوسط الحسابى الحقيقى  $\mu$  لدخل العائلة في مجتمع المدينة ، أو على أى دليل آخر يخص هذا المجتمع .

أما إذا اخترنا عينة من مجتمع هذه المدينة فإن التوزيع التكرارى لدخول العائلات في هذه العينة يسمى **بتوزيع العينة** . والوسط الحسابي  $\bar{x}$  لهذا التوزيع لا يكون عادة مساوياً للوسط الحسابي الحقيقي  $\mu$  للمجتمع ، إلا أننا قد نأخذ هذا الوسط الحسابي تحت شروط معينة ، كتقدير للوسط الحسابي الحقيقي للمجتمع . وبطبيعة الحال تختلف الأوساط الحسابية للعينات المأخوذة من نفس المجتمع حتى ولو كانت من نفس الحجم .

أما إذا فرضنا أننا حصلنا من المجتمع على جميع العينات التي من نفس الحجم  $n$  وأوجدنا الوسط الحسابي لكل من هذه العينات ثم وضعنا هذه الأوساط في توزيع تكرارى فإن هذا التوزيع يسمى **حيثخذ بتوزيع المعاينة للوسط الحسابي** ( أو للأوساط الحسابية ) للعينات ذوات الحجم  $n$  . ويمكننا أن نحسب الوسط الحسابي والانحراف المعيارى ومعامل الالتواء ... لهذا التوزيع . وبالمثل يمكن أن نتصور توزيع المعاينة للتيابن للعينات ذوات الحجم  $n$  أو توزيع المعاينة لأى مقياس آخر . في هذا المثال يستحيل عملياً إيجاد توزيع المعاينة للوسط الحسابي بالطريقة المذكورة لأن عدد العينات الممكنة هو عدد فلكى نعجز عن الحصول عليه . وفي المعتاد نحصل على توزيعات المعاينة بطرق رياضية ، إلا أنه لتوضيح مفهوم توزيع المعاينة نضرب المثال البسيط الآتي .

### مثال ( ٦ - ١ ) :

يتألف مجتمع من ٦ أرتاب أوزانها بالأوقيات ١١ ، ١٦ ، ١٢ ، ١٥ ، ١٦ ، ١٤ .

( أولاً ) أوجد توزيع المعاينة للوسط الحسابي للعينات ذوات الحجم ٢ ( اعتبر أن المعاينة مع الإرجاع ) .

( ثانياً ) أحسب الوسط الحسابى والتيابن للتوزيع الناتج وقارنهما بالوسط الحسابى والتيابن للمجتمع .

الحل :

(اولا ) بما أن المعاينة مع الإرجاع أى مع رد كل عدد يؤخذ إلى المجتمع قبل أخذ عدد آخر فإن عدد العينات ذوات الحجم ٢ هو  $6 \times 6 = 36$  لأن أى عدد من الأعداد الستة يمكن أن يقترن ( في عينة حجمها ٢ ) بأى عدد من الأعداد الستة ( بما في ذلك نفسه ) وتكون جميع العينات الممكنة هي :

(١١،١١) ، (١٦،١١) ، (١٥،١١) ، (١٢،١١) ، (١٦،١١) ، (١٤،١١)  
 (١١،١٦) ، (١٦،١٦) ، (١٥،١٦) ، (١٢،١٦) ، (١٦،١٦) ، (١٤،١٦)  
 (١١،١٢) ، (١٦،١٢) ، (١٥،١٢) ، (١٢،١٢) ، (١٦،١٢) ، (١٤،١٢)  
 (١١،١٥) ، (١٦،١٥) ، (١٥،١٥) ، (١٢،١٥) ، (١٦،١٥) ، (١٤،١٥)  
 (١١،١٦) ، (١٦،١٦) ، (١٥،١٦) ، (١٢،١٦) ، (١٦،١٦) ، (١٤،١٦)  
 (١١،١٤) ، (١٦،١٤) ، (١٥،١٤) ، (١٢،١٤) ، (١٦،١٤) ، (١٤،١٤)

الأوساط الحسابية لهذه العينات هي :

١٢,٥	،	١٣,٥	،	١٣	،	١١,٥	،	١٣,٥	،	١١
١٥	،	١٦	،	١٥,٥	،	١٤	،	١٦	،	١٣,٥
١٣	،	١٤	،	١٣,٥	،	١٢	،	١٤	،	١١,٥
١٤,٥	،	١٥,٥	،	١٥	،	١٣,٥	،	١٥,٥	،	١٣
١٥	،	١٦	،	١٥,٥	،	١٤	،	١٦	،	١٣,٥
١٤	،	١٥	،	١٤,٥	،	١٣	،	١٥	،	١٢,٥

ويمكن تلخيص هذه الأوساط في التوزيع التكرارى المبين بالعمودين الأول والثانى من الجدول (٦ - ١) وهذا هو توزيع المعاينة المطلوب .



المجدول (٦ - ١)

توزيع المعاينة للوسط الحسابي لأوزان ٦ أرباب للعينات من الحجم ٦

خطأ التقدير $k_i(\bar{x}_i - \mu)$	$k_i$	$\bar{x}_i$
٣ -	١	١١
٥ -	٢	١١,٥
٢ -	١	١٢
٣ -	٢	١٢,٥
٤ -	٤	١٣
٣ -	٦	١٣,٥
٠	٥	١٤
١	٢	١٤,٥
٥	٥	١٥
٦	٤	١٥,٥
٨	٤	١٦
صفر	٣٦	المجموع

(ثانيا) الوسط الحسابي لتوزيع المعاينة  $\bar{\mu} = \frac{1}{n} \sum k_i \bar{x}_i$

$$\text{أى أن } \bar{\mu} = \frac{1}{36} (16 \times 4 + \dots + 11,5 \times 2 + 11 \times 1)$$

$$= \frac{504}{36} = 14 \text{ أوقية}$$

$$\text{تباين توزيع المعاينة} = \frac{1}{p} [ \sum_{j=1}^p (x_j - \bar{x})^2 - \sum_{j=1}^p x_j^2 ]$$

$$\text{أى أن } \sigma^2 = \frac{1}{36} [ \frac{504}{36} - (16 \times 4 + \dots + 11,0 \times 2 + 11 \times 1) ]$$

$$\frac{11}{9} =$$

$$\text{الوسط الحسابى للمجتمع } \mu = \frac{1}{4} (14 + 16 + 10 + 12 + 16 + 11) = 14$$

$$\text{تباين المجتمع } \sigma^2 = \frac{1}{4} [ \frac{511}{4} - (14^2 + \dots + 16^2 + 11^2) ]$$

بالمقارنة نجد ما يلى :

$$\mu = \bar{\mu} \quad (\text{أ})$$

أى أن الوسط الحسابى لتوزيع المعاينة = الوسط الحسابى للمجتمع .

$$\frac{\sigma^2}{p} = \sigma^2 \quad (\text{ب})$$

أى أن تباين توزيع المعاينة = تباين المجتمع مقسوما على حجم العينة .

وهاتان النتيجةتان هما قاعدتان عامتان بالنسبة لتوزيع المعاينة للوسط الحسابى ويمكن إثباتهما رياضيا ، سواء كان المجتمع منتهيا والمعاينة مع الإرجاع أو كان لا نهائيا .

ملاحظة :

الأوساط الحسابية للعينات المأخوذة من مجتمع ما تختلف بطبيعة الحال عن بعضها البعض ، وحين نأخذ أحد هذه الأوساط  $\bar{x}$  لتقدير الوسط الحسابى  $\mu$  للمجتمع يكون هناك خطأ فى التقدير قدره  $\bar{x} - \mu$  ففى هذا المثال لدينا  $\mu = 14$  وإذا قدرنا هذا المتوسط من متوسط العينة الأولى وهو 11 يكون هناك خطأ قدره  $14 - 11 = 3$  وحين نقدره من متوسط العينة الثانية وهو 11,0

يكون هناك خطأ قدره  $2 = (11,5 - 14) \times 2 = 2,5 - 5$  (مع ملاحظة أن هناك عيتين متوسط كل منهما 11,5) وهكذا بالنسبة لأخطاء التقدير من متوسطات العينات الأخرى كما هو مبين بالعمود الثالث من الجدول (٦ - ١) ، حيث نلاحظ أيضا أن مجموع أخطاء التقدير هذه يساوى صفرا وهذا هو الذى أدى إلى المساواة (أ) .

#### A statistic

#### (٦ - ٢ - ١) الإحصاء

في المثال السابق كان لدينا ٣٦ وسطا حسابيا  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{36}$  وسمينا التوزيع التكرارى لهذه المتوسطات بتوزيع المعاينة للوسط الحسابى للعينات ذوات الحجم ٢. وحين نسحب عينة ما من مجتمع الأرانب فإننا لا نعلم مقدما المتوسط الذى نحصل عليه منها بل يكون ذلك متروكا للصدفة . ولذلك ننظر إلى هذه المتوسطات على أنها قيم مشاهدة من متغير عشوائى  $\bar{x}$  ونسمى هذا المتغير حينئذ بالإحصاءة  $\bar{x}$  ويكون التوزيع سابق الذكر هو توزيع المعاينة لهذه الإحصاءة .

كذلك إذا كانت  $\bar{x}_1, \bar{x}_2, \dots$  هي تباينات جميع العينات التى من نفس الحجم التى يمكن أن تؤخذ من مجتمع ما فإننا ننظر إليها كقيم متغير عشوائى  $\bar{s}^2$  ويكون توزيع هذه القيم هو توزيع المعاينة للإحصاءة  $\bar{s}^2$  . وبالمثل لأى مقياس إحصائى آخر .

#### Standard Error

#### (٦ - ٢ - ٢) الخطأ المعيارى

يسمى الانحراف المعيارى لتوزيع المعاينة لإحصاءة ما بالخطأ المعيارى لهذه الإحصاءة . ففي المثال السابق الخطأ المعيارى للوسط الحسابى أو للإحصاءة  $\bar{x}$  هو  $\sigma_{\bar{x}} = \sqrt{\frac{11}{36}} = 1,83$  تقريبا . والأخطاء المعيارية تلعب دورا رئيسيا فى نظرية العينات كما ستبين بعد .

نظرا لأن الوسط الحسابى  $\bar{x}$  للإحصاءة  $\bar{x}$  يساوى الوسط الحسابى  $\mu$  للمجتمع ، نقول إن  $\bar{x}$  هو مُقدَّر غير متحيز للوسط الحسابى للمجتمع unbiased estimator كما نقول للوسط الحسابى  $\bar{x}$  لأى عينة عشوائية بأنه تقدير غير متحيز لمتوسط المجتمع ، وأى فرق  $\bar{x} - \mu$  يعتبر كما سبق القول خطأً فى تقدير متوسط المجتمع من متوسط العينة ، ويقاس مدى دقة هذا التقدير بالخطأ المعياري  $\sigma_{\bar{x}}$  لتوزيع المعاينة للإحصاءة  $\bar{x}$  . لاحظ أن هذا الخطأ يتوقف على كل من  $\sigma$  ،  $n$  ويكون هذا الخطأ صغيرا إذا كانت  $\sigma$  صغيرة أو كان حجم العينة  $n$  كبيرا ومن ثم كان كبر العينة من العوامل الرئيسية لدقة التقدير .

وبصفة عامة إذا كنا نقدر أحد بارامترات مجتمع من عينة فإن التقدير يوصف بأنه غير متحيز إذا كان متوسط قيم هذا التقدير على جميع العينات العشوائية التى يمكن أخذها من المجتمع يساوى البارامتر الذى نقدره ، ومن الواضح أن هذه الصفة مطلوبة فى التقدير . وكما وجدنا أن  $\bar{x}$  تقدير غير متحيز للبارامتر  $\mu$  نجد أن  $E\left(\frac{1}{n-1} \sum (x_i - \bar{x})^2\right) = \sigma^2$  تقدير غير متحيز للتباين  $\sigma^2$  للمجتمع .

### (٦ - ٣) المعاينة من مجتمعات محدلة

#### Sampling From Normal Populations

إن الهدف الرئيسى من دراسة الإحصاء معرفة كيفية الحكم على المجتمعات وإصدار قرارات عنها عن طريق عينات مأخوذة منها ، وهذا ما سميناه بعملية الاستدلال الإحصائى . ولما كانت هذه العملية تتوقف على معرفتنا بتوزيعات الاحتمال للإحصاءات التى نستخدمها ، وجب علينا أن نخطط بهذه التوزيعات توطئة لتحقيق ذلك الهدف .

ومن أهم توزيعات المعاينة تلك التي تكون فيها المعاينة من مجتمعات معتدلة ،  
وستقدم في هذا البند عددا من هذه التوزيعات دون التعرض للبراهين الرياضية .

### (أولا) توزيع المعاينة للوسط الحسابي

#### Sampling Distribution of the Mean

إذا كان لدينا مجتمع معتدل وسطه الحسابي  $\mu$  وانحرافه المعياري  $\sigma$  فإن توزيع  
المعاينة للإحصاءة  $\bar{x}$  للعينات ذوات الحجم  $n$  المأخوذة من هذا المجتمع يكون  
توزيعا معتدلا وسطه الحسابي  $\mu$  وانحرافه المعياري  $\frac{\sigma}{\sqrt{n}}$  ( أى الخطأ المعياري ) ،  
وبالتالى فإن الإحصاءة

$$(1) \quad \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = z$$

يكون توزيعها مطابقاً للتوزيع المعتدل المعيارى . راجع الخاصة ( ذ ) بالبند  
( ٤ - ١ ) . والمفروض في هذه الإحصاءة أن تكون كل من  $\mu$  ،  $\sigma$  معروفة القيمة .  
كما أن الإحصاءة

$$(2) \quad \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = z$$

يكون توزيعها مطابقاً لتوزيع ت بدرجات حرية عددها  $n - 1$  . وتستخدم هذه  
الإحصاءة حينما تكون  $\sigma$  مجهولة القيمة وهذا ما يحدث في أغلب الحالات ،  
ونضطر حينئذ لاستخدام تباين العينة وهو

$$s^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right) \text{ كتقدير غير متحيز للتباين } \sigma^2 \text{ للمجتمع .}$$

(ثانياً) توزيع المعاينة للفرق بين وسطين حسابيين :

نفرض أن لدينا مجتمعين معتدلين وسطاهما الحسابيان  $\mu_1$  ،  $\mu_2$  وانحرافاهما المعياريان  $\sigma_1$  ،  $\sigma_2$  ونفرض أن الإحصاءة  $\bar{y}_1$  ترمز إلى الوسط الحسابي لعينة عشوائية حجمها  $n_1$  مأخوذة من المجتمع الأول وأن الإحصاءة  $\bar{y}_2$  ترمز إلى الوسط الحسابي لعينة عشوائية حجمها  $n_2$  مأخوذة من المجتمع الثاني . إذا كانت هاتان العيتان مستقلتين فإن توزيع المعاينة للإحصاءة  $(\bar{y}_1 - \bar{y}_2)$  يكون توزيعاً معتدلاً وسطه الحسابي  $(\mu_1 - \mu_2)$  وانحرافه المعياري  $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$  وبالتالي فإن الإحصاءة :

$$(3) \quad \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = z$$

يكون توزيعها مطابقاً للتوزيع المعتدل المعيارى .

والمفروض هنا أن تكون كل من  $\mu_1$  ،  $\mu_2$  ،  $\sigma_1$  ،  $\sigma_2$  معروفة القيمة .  
كما أن الإحصاءة

$$(4) \quad \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sqrt{c}$$

يكون توزيعها مطابقاً لتوزيع ت بدرجات حرية عددها  $n_1 + n_2 - 2$

$$(١) \quad \frac{{}_1^2 \bar{e}_n (1 - {}_1^2 u) + {}_2^2 \bar{e}_n (1 - {}_2^2 u)}{2 - {}_1^2 u + {}_2^2 u} = {}^2 \bar{e}_n \quad \text{حيث}$$

$$(١) \quad \frac{{}_1^2 (\bar{e}_n - {}_1^2 \bar{e}_n) + {}_2^2 (\bar{e}_n - {}_2^2 \bar{e}_n)}{2 - {}_1^2 u + {}_2^2 u} =$$

$$\text{وحيث } {}_1^2 \sigma = {}_2^2 \sigma$$

أما إذا كان  ${}_1^2 \sigma \neq {}_2^2 \sigma$  فستخدم الاحصاء

$$(٥) \quad \frac{({}_1^2 \mu - {}_2^2 \mu) - ({}_1^2 \bar{e}_n - {}_2^2 \bar{e}_n)}{\sqrt{{}_1^2 u / {}_1^2 \bar{e}_n + {}_2^2 u / {}_2^2 \bar{e}_n}} = \bar{e}_n$$

التي لها توزيع ت بدرجات حرية  $\left( \frac{{}_1^2 \bar{e}_n}{{}_1^2 u} + \frac{{}_2^2 \bar{e}_n}{{}_2^2 u} \right) \div$

$$\left[ (1 - {}_1^2 u) \div \left( \frac{{}_1^2 \bar{e}_n}{{}_1^2 u} \right) + (1 - {}_2^2 u) \div \left( \frac{{}_2^2 \bar{e}_n}{{}_2^2 u} \right) \right]$$

ملاحظة :

في المعاينة من أى مجمع ( ليس مختللاً بالضرورة ) نستطيع رياضياً إثبات ما  
يل اعتماداً على النظرية الشهيرة المسماة بنظرية النهاية المركزية central limit  
theorem .

( أ ) يقترب توزيع الإحصاء  $\bar{e}_n$  المعرفة في (١) من التوزيع المعتدل المعيارى  
حين يقترب حجم العينة من اللانهاية .

(ب) يقترب توزيع الإحصاءة  $\bar{y}$  المعرفة في (٣) من التوزيع المعتدل المعياري حين تقترب كل من  $n_1$  ،  $n_2$  من اللانهاية .

إن المعنى التطبيقي لهاتين النظريتين أنه عند المعاينة من مجتمع غير معتدل وبشرط أن تكون العينات كبيرة الحجم ( أكبر من ٣٠ ) يجوز عملياً اعتبار أن توزيع كل من  $\bar{y}_1$  ،  $\bar{y}_2$  هو بالتقريب توزيع معتدل معياري ، وكلما كبر حجم العينات كلما قل الخطأ الناشئ عن هذا التقريب .

( ثالثاً ) توزيع المعاينة لخارج قسمة تباينين :

إذا كانت الإحصاءتان  $\bar{y}_1$  ،  $\bar{y}_2$  ترمزان إلى تباينين عيشتين عشوائيتين مستقلتين مأخوذتين من مجتمعين معتدلين لهما نفس التباين ( $\sigma^2 = \sigma^2$ ) فإن الإحصاءة

$$\frac{\bar{y}_1^2}{\bar{y}_2^2} = \bar{y} \quad (٦)$$

يكون توزيعها مطابقاً لتوزيع ف بدرجتي حرية  $n_1 - 1$  ،  $n_2 - 1$  حيث  $n_1$  ،  $n_2$  هما حجم العينتين . ويلاحظ أنه إذا كانت  $\bar{y}_1$  ،  $\bar{y}_2$  قيمتين مأخوذتين من عيشتين عشوائيتين مستقلتين من مجتمع واحد أو من مجتمعين لهما نفس التباين فإن النسبة  $\bar{y}_1^2 / \bar{y}_2^2$  تكون قريبة من الواحد لأن كلا من  $\bar{y}_1$  ،  $\bar{y}_2$  تقديراً لنفس التباين  $\sigma^2$  .

(٦ - ٤) المعاينة من توزيع ذى الحدين :

#### SAMPLING FROM A BINOMIAL DISTRIBUTION

اعتبر مجتمعاً مقسماً إلى قسمين منفصلين من حيث وقوع أو عدم وقوع حدث معين أ . افرض أن احتمال وقوع هذا الحدث في أى تجربة واحدة هو مقدار ثابت ح واحتمال عدم وقوعه ك = ١ - ح . حسب البند (٣ - ٣) ونمت الشروط



المذكورة يكون توزيع عدد مرات وقوع الحدث في  $n$  من التجارب هو توزيع ذي الحدين ذيلاه  $n$  ،  $h$  ، وسطه الحسابي  $nh$  وانحرافه المعياري  $\sqrt{nh}$   $h$  ويكون التكرار النسبي لوقوع هذا الحدث هو خارج قسمة عدد مرات وقوع الحدث على العدد  $n$  وعلى ذلك فإن متوسط هذه النسبة يساوي :

$$\frac{h}{n} = \frac{nh}{n} \quad \text{وانحرافها المعياري} \quad \sqrt{\frac{nh}{n}} = \sqrt{h}$$

وإذا أخذنا من هذا المجتمع جميع العينات ذوات الحجم  $n$  وحسبنا في كل منها القيمة  $\bar{h}$  لعدد مرات وقوع ذلك الحدث ثم حسبنا نسبة هذا العدد وهو  $r = \frac{\bar{h}}{n}$  فإننا نحصل على توزيع المعاينة لهذه النسبة .

وقد رأينا في البند (٤ - ٦) أنه إذا كانت  $n$  كبيرة ولم تكن أى من  $h$  أو  $nh$  قريبة من الصفر فإن :

(أ) توزيع الاحتمال لعدد مرات وقوع الحدث في العينات ذوات الحجم  $n$  يقترب من توزيع معتدل وسطه الحسابي  $nh$  ، وانحرافه المعياري  $\sqrt{nh}$   $h$  وبالتالى فإن توزيع الإحصاءة

$$(٧) \quad \frac{\bar{h} - nh}{\sqrt{nh}} = \bar{v}$$

يقترب من التوزيع المعتدل المعياري مع (١ ، ٠) .

(ب) وبالمثل فإن توزيع الإحصاءة :

$$(٨) \quad \frac{\bar{h} - nh}{\sqrt{\frac{nh}{n}}} = \bar{v}$$

يقترب من التوزيع المعتدل المعياري .

## ٦ - ٥) اختبارات الفروض : TESTS OF HYPOTHESES

**الفرض الإحصائي :** هو جملة أو مقولة نذكرها عن مجتمع أو عدة مجتمعات بهدف اختبار صواب أو خطأ هذه الجملة ، ولمعرفة هذا الصواب أو هذا الخطأ نستعين بما يسمى باختبارات الفروض وهى اختبارات تبني على إحصاءات تكون توزيعاتها معروفة لنا مثل تلك التي وردت في البندين (٦ - ٣) و (٦ - ٤) السابقين .

**واختبار الفرض** هو بكل بساطة قاعدة تؤدي إلى اتخاذ قرار برفض أو قبول الفرض فور حصولنا على قيم مشاهدة في عينة عشوائية . وفي المعتاد نضع فرضاً نرمز له بالرمز  $F$  . نسميه **الفرض الصفري** أو بفرض العدم  $null hypothesis$  نصوغه بحيث يعبر عن انعدام الفرق بين شيئين ، مثلاً  $\mu_1 = \mu_2$  (أى  $\mu_1 - \mu_2 = 0$ ) أو  $\mu = 1$  (أى  $\mu - 1 = 0$ ) أو لا توجد علاقة بين متغيرين . وأى فرض يراد اختباره ضد الفرض الصفري يسمى بالفرض الآخر أو بالفرض البديل  $alternative hypothesis$  ويرمز له بالرمز  $F$  وقد يأخذ هذا الفرض الشكل .

$$\mu_1 \neq \mu_2 \text{ أو } \mu_1 < \mu_2 \text{ أو } \mu_1 > \mu_2 .$$

ولاختبار الفرض الصفري  $F$  ضد الفرض الآخر  $F$  نبدأ باختيار إحصاءة تناسب الفرض المختبر بشرط أن يكون توزيع المعاينة لهذه الإحصاءة معلوماً لنا ثم نحدد في هذا التوزيع منطقة  $M$  يكون احتمال وقوع القيم المشاهدة لهذه الإحصاءة فيها هو احتمال صغير  $\alpha$  ، مثلاً  $0.05$  ،  $0.01$  ،  $0.001$  ، ... وتسمى هذه المنطقة بالمنطقة الحرجة أو بمنطقة الرفض  $critical region$  or  $region of rejection$  أما الاحتمال  $\alpha$  فيسمى بمستوى الدلالة للاختبار  $level of significance$  وتأخذ القاعدة التي يعطها الاختبار الصورة الآتية :

« إذا وقعت قيمة مشاهدة لهذه الإحصاءة ، محسوبة على أساس صحة الفرض الصفري  $F$  ، داخل المنطقة الحرجة نرفض  $F$  عند مستوى الدلالة  $\alpha$  وإلا نقبل  $F$  . »

وذلك على أساس أن احتمال وقوع القيمة المشاهدة داخل المنطقة الحرجة هو احتمال ضئيل يجعلنا نشك في صحة الفرض الصفري . ونظراً لأن القيم المشاهدة يمكن أن تقع في أي جزء من التوزيع فإن هذه القاعدة تعني أننا نغامر برفض الفرض الصفري ( عند المستوى  $\alpha$  ) مع علمنا بأننا قد نكون مخطئين في رفضنا هذا ، وإن كان احتمال هذا الخطأ هو احتمال صغير لا يزيد عن  $\alpha$  . أما إذا وقعت القيمة المشاهدة في المنطقة المكملة للمنطقة الحرجة ، وتسمى بمنطقة القبول ، فلا يسعنا إلا قبول الفرض ف. على أساس أن احتمال وقوع القيمة المشاهدة فيها هو احتمال كبير  $1 - \alpha$  .

وجدير بالملاحظة أنه بينما يمكننا الاختبار من هدم الفرض الصفري إلا أنه لا يستطيع أن يثبت صحته وكل ما يستطيع أن يفعله لصالح هذا الفرض هو أن يبين عدم وجود ما يتعارض معه من واقع العينة المستخدمة ، وحين نقول إننا نقبل ف. لا نعني أننا أثبتنا صحته وإنما نعني أن ما لدينا من بيانات لا تعطي دليلاً كافياً يدعو إلى رفضه ، وحين نقول إننا نرفض ف. فإننا لا نعني أننا أثبتنا زيفه وإنما نعني أن ما لدينا من بيانات لا تدعم صحته بل تشير إلى صحة الفرض الآخر ف.

واختبار الفرض ف. عن مجتمع ضد فرض آخر ف. يستلزم بطبيعة الحال الحصول على عينة عشوائية من هذا المجتمع ، ويفضل اختيار قيمة  $\alpha$  حتي قبل اختيار العينة . ويتوقف تحديدنا لقيمة  $\alpha$  على طبيعة المشكلة التي نتناولها ودرجة المغامرة التي نقبلها لتحمل مسؤولية الخطأ المحتمل . وبصفة عامة يمكن تلخيص خطوات إجراء اختبار الفروض فيما يلي :

( أولاً ) نحدد الفرض الصفري ف. والفرض الآخر ف. من واقع المشكلة التي نتناولها .

( ثانياً ) نحدد الإحصاءة التي تناسب الفرض المختبر ونحدد المنطقة الحرجة م في توزيع هذه الإحصاءة بناء على مستوى الدلالة  $\alpha$  السابق اختياره .

( ثالثاً ) بحسب قيمة الإحصاءة من واقع البيانات المشاهدة في عينة عشوائية وعلى أساس أن الفرض الصفري صحيح .

( رابعاً ) الاستنتاج : نتخذ قراراً برفض أو قبول الفرض الصفري ف عند مستوى الدلالة  $\alpha$  بحسب وقوع القيمة المحسوبة للإحصاءة داخل أو خارج المنطقة الحرجة .

ويجدر الإشارة هنا إلى أن اختبارات الفروض ما هي إلا وسيلة حسابية تستخدم البيانات التي حصلنا عليها من العينات لإلقاء الضوء على صحة أو خطأ الفرض الصفري ، وينبغي للباحث عند إصدار قراره أن يضيف إلى نتيجة الاختبار جميع ما لديه من معلومات وخبرات وأبحاث سابقة عن موضوع الدراسة .

كما يجدر الإشارة إلى أن حجم العينة يلعب دوراً هاماً في عملية الاستدلال الاحصائي . وحين تكون العينة صغيرة فإن اختبار الدلالة لا يؤدي إلى رفض الفرض الصفري إلا إذا كان هذا الفرض خاطئاً بدرجة كبيرة ، وبالعكس حين تكون العينة كبيرة فإن أي انحراف صغير عن الفرض الصفري يسهل اكتشافه كإنحراف ذي دلالة . وبصفة عامة يفضل أن يكون حجم العينة كبيراً بدرجة كافية منعا للوقوع فيما يسمى بالخطأ من النوع الثاني ، وهذا ما سوف نتناوله في الفصل السابع .

في البنود ( ٦ - ٦ ) ، ( ٦ - ٧ ) ( ٦ - ٨ ) الآتية نتناول ثلاثة من أشهر اختبارات الفروض تعرف باختبار ت واختبار  $\chi^2$  ( كاً ) واختبار ف .

## THE t-TEST ( ٦ - ٦ ) اختبار ت :

هناك عدة إحصاءات تطابق توزيعاتها توزيع ت السابق دراسته . وإذا بني اختبار فروض على أي من هذه الإحصاءات قبل إنه اختبار ت . والصورة العامة لهذه الإحصاءات هي :

ت = القيمة المشاهدة للإحصاءة - الوسط الحسابي للإحصاءة  
تقدير للخطأ المعياري للإحصاءة

بشرط أن يكون للإحصاءة توزيع معتدل وأن تكون القيمة المشاهدة هي تقدير غير متحيز لوسط الإحصاءة .

(٦ - ٦ - ١) اختبار فرض عن الوسط الحسابي لمجتمع معتدل :

مثال (٦ - ٢) :

تقضي التعليمات الحكومية بأن تكون الجرعة القياسية من مستحضر بيولوجي ٦٠٠ وحدة نشاط في السيتيمتر المكعب . اختبرت عينة عشوائية حجمها ١٠ من إنتاج شركة ما ووجد أن متوسطها ٥٩٢,٥ وحدة نشاط في السيتيمتر المكعب بانحراف معياري ١١,٢ وحدة . هل نستطيع القول بأن إنتاج هذه الشركة يتماشى مع التعليمات الحكومية ؟

الحل :

نريد أن نختبر ما إذا كان الوسط الحسابي  $\mu$  لمجتمع المستحضر الذي تنتجه الشركة يساوي الجرعة القياسية التي حددتها التعليمات الحكومية وهي ٦٠٠ وحدة نشاط /  $cm^3$  . تتبع الخطوات الأربع المشار إليها في البند السابق .

(١) الفرض الصفري  $F_0$  :  $\mu = 600$

الفرض الآخر  $F_1$  :  $\mu \neq 600$

$\alpha = 0,05$  ،

(٢) لنهتدى إلى الإحصاءة المناسبة نذكر أننا نريد اختبار فرض عن قيمة الوسط الحسابي  $\mu$  للمجتمع عن طريق قيمة الوسط الحسابي  $\bar{x}$  لعينة وتوقع أنه إذا كان هذا الفرض صحيحاً فإن  $\bar{x}$  تكون قريبة من  $\mu$  ، أي أننا نريد اختبار ما إذا كان الفرق بين  $\bar{x}$  و  $\mu$  هو فرق صغير نعزوه إلى الصدفة أو فرق جوهري ( ذو دلالة

بدعونا إلى عدم الثقة في القيمة المفروضة للدليل  $\mu$  . ( هذا على أساس أن العينة هي عينة عشوائية ممثلة للمجتمع تمثيلاً جيداً وبالتالي فإن وسطها الحسابي  $\bar{x}$  هو عدد نثق فيه ) . هذا التحليل يشير مباشرة إلى أن أنسب إحصاءة لقياس صغر أو كبر هذا الفرق هي الإحصاءة (٢) وهي :

$$T = \frac{\bar{x} - \mu}{\sqrt{s^2/n}}$$

التي نعلم أن توزيعها يطابق توزيع  $T$  بدرجات حرية عددها  $\nu = n - 1$  ، مع ملاحظة أن هذه الإحصاءة تتعلم عندما تتساوى قيمتي  $\mu$  ،  $\bar{x}$  وتكبر قيمتها المطلقة كلما كبر الفرق بينهما .

$$\{T : |T| < t_{\alpha/2, \nu}\}$$



الشكل (١-٩)

وهي تتألف من جزئين واقعيين أسفل جانبي منحنى  $T$  واحتمال كل منهما  $\alpha/2$  . وبالتالي فإن احتمال وقوع قيمة الإحصاءة في هذه المنطقة هو احتمال صغير  $\alpha$  . (٣) نحسب قيمة هذه الإحصاءة من بيانات العينة وعلى أساس صحة الفرض الصفرى . وإذا رمزنا لهذه القيمة بالرمز  $T$  نجد أن :

$$T = \frac{6.00 - 5.92, 5}{\sqrt{1.0} / 11, 2} = 2, 12 \text{ حيث } \nu = 10 = 11 - 1$$

فإذا كنا قد اخترنا  $\alpha = 0.05$  نجد من جدول ت أن القيمة الحرجة التي تحدد منطقة الرفض هي :  $t_{0.05, 19} = 1.729$

(٤) الاستنتاج : بما أن  $t = -2.12$  لا تقع في المنطقة الحرجة لا يكون لدينا دليل ضد الفرض الصفري عند مستوى الدلالة  $0.05$  أى أننا نقبل الفرض أن  $\mu = 600$  ونقرر بأن إنتاج الشركة يتمشى مع التعليمات الحكومية .

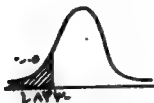
الاختبار ذو الجانب الواحد وذو الجانبين :

في المثال السابق اخترنا الفرض الصفري  $F : \mu = 600$  ضد الفرض  $F_1 : \mu \neq 600$  وعلى ذلك كان لنا أن نرفض  $F$  في حالتين هما : أن تكون  $\mu > 600$  أو تكون  $\mu < 600$  ولذلك فملت المنطقة الحرجة كلا جانبي توزيع  $t$  . ونقول حينئذ إن الاختبار ذو جانبيين أو إنه اختبار غير اتجاهي .

ولكن إذا كانت زيادة وحدات النشاط في المستحضر البيولوجي عن المستوى القياسي لا ينتج عنها ضرر بيننا نقصان عنه يفقد المستحضر فعاليته فإن اهتمامنا يكون منصفاً على ما إذا كان متوسط العينة أصغر صغراً ذا دلالة من المستوى القياسي . في هذه الحالة يكون المطلوب اختبار الفرض الصفري  $F : \mu = 600$  ضد الفرض  $F_1 : \mu > 600$  وتكون المنطقة الحرجة في جانب واحد فقط من التوزيع هو الجانب الأيسر ، أى أن المنطقة الحرجة تكون في هذه الحالة على الصورة ،

$$\{t : t > t_{1-\alpha, n-1}\} = \alpha$$

→ منطقة القبول ←



العكس (١-٢)

ونقول حيث إن الاختبار ذو جانب واحد أو إنه اختبار اتجاهي . فإذا كان مستوى الدلالة هو  $\alpha = 0.05$  فإن القيمة التي تحد المنطقة الحرجة من اليمين وتجعل المساحة بينها وبين المنحنى  $0.05$  يمكن إيجادها من جدول ت باستخدام الاحتمال  $\alpha = 0.10$  .

من الجدول نجد أن ت  $1.833 =$  [١١.٠١]

إذن القيمة التي تحد المنطقة الحرجة من اليمين هي  $1.833$  .

بما أن  $1.833 > 2.12$  فإن القيمة المشاهدة  $2.12$  تقع في المنطقة الحرجة وبالتالي نرفض الفرض الصفري ف عند مستوى الدلالة  $0.05$  لصالح الفرض الآخر ونقرر أن متوسط إنتاج الشركة يقل عن المستوى الذى تحدده التعليمات الحكومية .

ولا نعجب من اختلاف هذه النتيجة عن النتيجة السابقة لأن كلا منهما يجيب على سؤال مختلف هو السؤال الذى تتطلب المشكلة الإجابة عنه .

وبالمثل يمكن أن يكون اهتمامنا منصّباً على الجانب الأيمن فقط من التوزيع حيث تكون المنطقة الحرجة على الصورة  $2 = \{ ت : ت < ٢.١٢ \}$  . وفي تناول أى مشكلة علينا أن نفكر جيداً قبل أن نحدد ما إذا كانت تتطلب اختباراً ذا جانب واحد أو ذا جانبيين تحسباً من الوقوع في خطأ في عملية الاستدلال . وهذا التحديد ينبغى أن يتقرر عند تصميم التجربة وقبل جمع البيانات وبحسب التساؤل الذى تطرحه المشكلة . فمثلاً إذا كنا نقارن أداء مجموعة من الطلاب بمستوى معروف فإن اهتمامنا ينصب على معرفة ما إذا كانت المجموعة أحسن أو أسوأ من هذا المستوى ، وهنا يجب أن يكون الاختبار ذا جانبيين . أما إذا كنا نقارن نوعاً جديداً من الدواء بنوع تقليدى فإن اهتمامنا يكون منصّباً على معرفة ما إذا كان الدواء الجديد أفضل من التقليدى وهنا يجب أن يكون الاختبار ذا جانب واحد .



## ملاحظة (١) :

يُطلب استخدام اختبارات بهذه الصورة تحقق الشرطين الآتيين :

- (أ) أن تكون العينة عشوائية بسيطة .
- (ب) أن يكون المجتمع معتدلاً (أو يمكن اعتباره معتدلاً) .

## (٦ - ٢ - ٢) فترات الثقة للوسط الحسابي لمجتمع معتدل :

### CONFIDENCE INTERVALS

إذا كان الوسط الحسابي  $\mu$  لمجتمع ما مجهولاً ، يمكن تقديره بواسطة الوسط الحسابي  $\bar{x}$  لعينة عشوائية ذات حجم مناسب مأخوذة من هذا المجتمع ، ومثل هذا التقدير يسمى بالتقدير بنقطة point estimation ، ولكن نظراً لأن متوسطات العينات المأخوذة من نفس المجتمع تختلف من عينة إلى أخرى حتى ولو كانت العينات من نفس الحجم فإن هذا التقدير يشوبه بعض الصعوبات خاصة في تحديد مدى الثقة التي نضعها فيه .

ولذلك يفضل في كثير من الأحيان تقدير الوسط الحسابي وغيره من أدلة المجتمع عن طريق ما يسمى بالتقدير بفترة interval estimation حيث تحدد فترة يقع فيها الدليل المجهول بدرجة ثقة معينة .

حينما يكون المجتمع معتدلاً وسحبنا منه عينة عشوائية بسيطة حجمها  $n$  ووسطها الحسابي  $\bar{x}$  وانحرافها المعياري  $\sigma$  فإن الفترة :

$$(\bar{x} - \frac{\sigma}{\sqrt{n}} t_{1-\alpha/2} , \bar{x} + \frac{\sigma}{\sqrt{n}} t_{1-\alpha/2}) \quad (١٠)$$

تسمى بفترة ثقة بدرجة  $(1 - \alpha)$  لمتوسط المجتمع  $\mu$  . وهذه العبارة لا تعني أن احتمال وقوع  $\mu$  في هذه الفترة هو  $(1 - \alpha)$  لأن  $\mu$  عدد ثابت لا توزيع له وإنما تفسر هذه العبارة كما يلي :

« إذا مسحنا جميع العينات التي من نفس الحجم وأوجدنا فترة الثقة بدرجة  $(1 - \alpha)$  لكل منها فإن  $(1 - \alpha)$  من هذه الفترات تشمل البارامتر  $\mu$  » ، وهذا ما يمكن إثباته رياضياً .

ويسمى العددان  $\bar{x} \pm \frac{s}{\sqrt{n}}$  اللذان يحذان فترة الثقة بحدى الثقة

بدرجة  $(1 - \alpha)$  أما العدد  $\frac{s}{\sqrt{n}}$  فهو تقدير للخطأ المعياري للإحصاءة  $\bar{x}$  ،

كما تسمى  $\alpha$  بمعامل الثقة .

مثال (٦ - ٣) :

أوجد فترة ثقة بدرجة ٩٥٪ للوسط الحسابي  $\mu$  لمجتمع المستحضر البيولوجي المذكور في المثال (٦ - ٢) من واقع البيانات المعطاة في هذا المثال .

الحل :

لدينا  $n = 10$  ،  $\bar{x} = 592,5$  ،  $s = 11,2$  ،  $1 - \alpha = 0,95$  ومنها  $\alpha = 0,05$

$$t_{(n-1), \alpha/2} = 2,262 \text{ ، } \frac{s}{\sqrt{n}} = \frac{11,2}{\sqrt{10}} = 3,54$$

بالتعويض في (١٠) نجد أن :

$$\text{الحد الأدنى للفترة} = 592,5 - 2,262 \times 3,54 = 584,49$$

الحد الأعلى للفترة =  $592,5 + 3,54 \times 2,262 = 600,51$   
 إذن الفترة (584,49 ، 600,51) هي فترة ثقة بدرجة 95٪ لتوسط المجتمع .

ملاحظة (٢) :

في المثال (٦ - ٢) قبلنا الفرض أن  $\mu = 600$  ضد الفرض  $\mu \neq 600$  عند مستوى الدلالة 0,05 ، ونلاحظ أن العدد 600 يقع في فترة الثقة التي أوجدناها في المثال (٦ - ٣) ، والواقع أن أى عدد نفضه عن متوسط المجتمع  $\mu$  يكون مقبولا عند المستوى 0,05 طالما كان واقعاً في هذه الفترة . تحقق من ذلك باختيار بعض القيم المناسبة واختبارها كما بالبند (٦ - ٦ - ١) .

(٦ - ٦ - ٣) مقارنة متوسطي مجتمعين معتلين :

(أو اختبار دلالة الفرق بين متوسطي عيتين مستقلتين مأخوذتين من مجتمعين معتلين) .

مثال (٦ - ٤) :

في دراسة عن نبات حنك السبع كان المطلوب مقارنة الوسط الحسابي للنباتات ذات الزهور الحمراء من النوع Suttons Eclipse بالوسط الحسابي للنباتات ذات الزهور البيضاء من النوع Suttons Intermediate White وقد نتجت البيانات الآتية من عيتين عشوائيتين مستقلتين .

ذات الزهور الحمراء :  $\bar{x} = 43$  ،  $\bar{s}^2 = 134,76$  ،  $n = 32$  ،  $\bar{x} = 32,35$

ذات الزهور البيضاء :  $\bar{x} = 21$  ،  $\bar{s}^2 = 169,50$  ،  $n = 22$  ،  $\bar{x} = 22,93$

اختبر ما إذا كان الفرق المشاهد بين متوسطي العيتين يرجع إلى تقلبات العينات أو إلى وجود فرق حقيقي بين المتوسطين  $\mu_1$  ،  $\mu_2$  للمجتمعين اللذين سحبت منهما العينتان . اقرص اعتدابه المحتملين ونساوي تباينهما وحذ  $\alpha = 0,05$  .

الحل :

الفرض الصفري ف:  $\mu_1 = \mu_2$  (لا يوجد فرق بين متوسطي المجتمعين)  
 الفرض الآخر ف:  $\mu_1 \neq \mu_2$  (اختبار ذو جانبيين)  
 الإحصاء المناسبة هنا هي الإحصاءة (٤) وهي :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{ع}$$

التي نعلم أن توزيعها يطابق توزيع ت بـ درجات حرية عددها  $v = n_1 + n_2 - 2$

$$\frac{{}^2\text{ع} (1 - \alpha) + {}^1\text{ع} (1 - \alpha)}{v} = {}^2\text{ع}$$

على فرض أن تبايني المجتمعين متساويان

$$\{t : |t| < t_{\alpha/2, v}\}$$

نحسب ت<sub>٢</sub> من بيانات العينة وعلى أساس صحة الفرض الصفري .

$${}^2\text{ع} = \frac{{}^1\text{ع} \times 20 + {}^2\text{ع} \times 42}{62} = 1986,4135$$

$$1986,4135 =$$

$$\therefore \text{ع} = 44,569$$

$$\text{تقدير الخطأ المعياري} = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \times 44,569 = \sqrt{\frac{1}{21} + \frac{1}{43}} \times 44,569$$

$$\therefore \text{ت} = \frac{169,50 - 134,76}{11,800} = 2,93 \quad \text{حيث } v = 21 + 43 - 2 = 62$$

من جدول ت نجد أن ت<sub>٠,٠٥,٦٢,٠٠٠</sub>

بما أن  $2.93 - > 2$  فإن القيمة ت  $= 2.93$  تقع في المنطقة الحرجة ( في الجانب الأيسر ) وإذن نرفض الفرض الصفري ف :  $\mu_1 = \mu_2$  عند مستوى الدلالة 0.05 ونقرر أن العيتين مأخوذتان من مجتمعين مختلفين في وسطهما الحسابي .

### ملاحظة (٣)

يتطلب استخدام اختبار ت بهذه الصورة تحقق الشروط الآتية :

- (أ) أن تكون كل من العيتين عشوائية بسيطة .
  - (ب) أن تكون كل من العيتين مأخوذة من مجتمع معتدل .
  - (ج) أن تكون العيتان مستقلتين .
  - (د) أن يكون تباين المجتمعين متساويان .
- إذا لم تكن العيتان مستقلتين فإن اختبار ت للمقارنة بين متوسطي المجتمعين يأخذ الصورة التي سترد في البند (٨ - ٧ - ١) القادم .

فترات الثقة للفرق بين متوسطي مجتمعين محتلين :

يمكن إثبات أنه تحت الشروط المذكورة ، يصلح العدنان :

$$(11) \quad (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

كحدين لفترة ثقة بدرجة (١ -  $\alpha$ ) للفرق ( $\mu_1 - \mu_2$ ) بين متوسطي المجتمعين ،

حيث  $z_{\alpha/2}$  = ع  $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$  هو تقديره للخطأ المعياري للإحصاءة  $\bar{x}_1 - \bar{x}_2$ .

ففي المثال (٦ - ٤) نجد أن :

$$\begin{aligned} \text{الحد الأدنى لفترة الثقة} &= (134,76 - 169,5) = 2 \times 11,855 - 11,03 = 11,03 \\ \text{الحد الأعلى لفترة الثقة} &= (134,76 - 169,5) + 2 \times 11,855 = 58,45 \\ \text{أى أننا يمكن أن نأخذ الفترة } &(11,03, 58,45) \text{ كفترة ثقة بدرجة } 95\% \text{ للفرق} \\ &\text{بين متوسطى المجموعين .} \end{aligned}$$

### تمارين (٦ - ١)

في المسائل الآتية اعتبر أن المجموعات معطلة وأن العينات عشوائية ومستقلة وأن مستوى الدلالة  $\alpha = 0,05$  ما لم يذكر غير ذلك .

(١) وزنت ١٩ علراء جرادة فكانت الأوزان كما على بالجرامات :

$$\begin{array}{cccccccccccccccc} 0,70 & 0,52 & 0,68 & 0,93 & 0,82 & 1,22 & 1,05 & 1,18 & 0,95 & 0,67 & 0,84 & 0,89 & 1,12 & 0,54 & 0,67 & 0,92 & 0,35 & 0,59 & 0,57 \end{array}$$

(أ) اختبر الفرض أن الوسط الحسابي للوزن في مجتمع هذه الحشرة هو

$$\mu = 1 \text{ ضد كل من الفرضين الآتين (١) } \mu \neq 1, \text{ (٢) } \mu > 1$$

(ب) أوجد فترة ثقة بدرجة 95% للدليل  $\mu$  .

(٢) باستخدام عينة حجمها  $n = 15$  ووسطها الحسابي  $\bar{x} = 85,3$  وتباينها

$$s^2 = 13 \text{ اختبر الفرض أن متوسط المجتمع } \mu = 80 \text{ وأوجد فترة ثقة بدرجة } 99\% \text{ لهذا المتوسط .}$$

(٣) في عينة عشوائية من ٢٠ شخصاً يعالجون من مرض معين وجد أن عدد

الكريات الدموية البيضاء بقدرة بالآلاف كالآتي . أوجد فترة ثقة بدرجة 95% لمتوسط عدد الكريات البيضاء في مجتمع هؤلاء المرضى

$$\begin{array}{cccccccccccccccc} 8 & 7 & 10 & 9 & 7 & 12 & 4 & 6 & 13 & 8 & 7 & 13 & 8 & 11 & 5 & 10 & 13 & 12 & 10 \end{array}$$

(٤) في بحث ثبات reliability إحدى طرق القياس في تجارب التبريد وجدت القيم الآتية لدرجات الحرارة في عيتين أ ، ب :

أ : ١٠٦,٩ ١٠٦,٣ ١٠٧,٠ ١٠٦,٠ ١٠٤,٩  
ب : ١٠٦,٥ ١٠٦,٧ ١٠٦,٨ ١٠٦,١ ١٠٥,٦

هل الفرق بين متوسطي العيتين ذو دلالة عند المستوى ٠,٠٥ ؟  
( لتسهيل حساب التباين اطرح ١٠٠ من جميع الأعداد ) .

(٥) أخذت عيتان متكافئتان من الأبقار ووضعتا تحت نفس الظروف فيما عدا أنهما أعطيتا نوعين مختلفين من الغذاء ، فكانت الزيادة في الأوزان بالأرطال بعد شهرين كما يلي :

العينة الأولى : ٣٣ ٦٦ ٢٦ ٤٣ ٤٦ ٥٥ ٥٤  
العينة الثانية : ٥٣ ٥٣ ٣٧ ٧٣ ٥٨ ٦١ ٣٨

( أ ) هل يمكن القول بأن نوعي الغذاء لا يختلفان في متوسط الزيادة في وزن الأبقار ؟

(ب) أوجد فترة ثقة بدرجة ٩٥٪ للفرق بين متوسطي المجتمعين .

(٦) طرح فرض بأن متوسط عدد الزهور الشعاعية ray florets في زهرة ما هو  $\mu = ٥٠$  . أخذت عينة حجمها ٥١ من هذه الزهرة فوجد أن متوسط عدد الزهور الشعاعية ٤٤,٣٣ بانحراف معياري ١,٦٧

( أ ) اختبر الفرض أن  $\mu = ٥٠$

(ب) اختبر الفرض أن  $\mu > ٥٠$

(ج) أوجد فترة ثقة بدرجة ٩٥٪ للدليل  $\mu$

( د ) أوجد فترة ثقة بدرجة ٩٩٪ للدليل  $\mu$

## (٦ - ٦ - ٤) اختبار استقلال الأحداث النادرة :

نعلم أنه في توزيع بواسون يتساوى التباين والوسط الحسابي أى أن النسبة بينهما تساوى الواحد الصحيح - راجع الخاصة (٩) من البند (٣ - ٤) . وفي تناولنا للأحداث النادرة في البندين (٣ - ٤ - ٢) ، (٣ - ٤ - ٣) علمنا أن الأحداث النادرة تتوزع بواسونياً بشرط أن تقع مستقلة عن بعضها بمعنى أن يكون وقوع هذه الأحداث غير متأثر إلا بالعوامل العشوائية وحدها . وفي كثير من التطبيقات يكون من المرغوب فيه الكشف عن هذه الاستقلالية ثم بحث ما قد يوجد من أنماط نتيجة لعدم توفر هذه الخاصة - راجع البند (٣ - ٤ - ٣) . ولتحقيق هذا الفرض نحصل على عينة عشوائية من تلك الأحداث وليكن حجمها  $n$  ووسطها الحسابي  $\bar{x}$  وتباينها  $s^2$  ونضع الفرض الصفري أن هذه الأحداث تقع عشوائياً مستقلة عن بعضها وبالتالي تتوزع بواسونياً وتكون النسبة بين التباين والوسط الحسابي في المجتمع هي  $1 = \frac{s^2}{\bar{x}}$  فإذا كان هذا الفرض صحيحاً فإن النسبة  $\bar{x} = s^2 / \bar{x}$  / الملاحظة في العينة ينبغي أن تكون قريبة من الواحد الصحيح ، أما إذا كان هذا الفرض خاطئاً فإن الفرق بين هذه النسبة والواحد يكون فرقاً ذا دلالة .

ولاختبار صحة الفرض  $1 = \frac{s^2}{\bar{x}}$  نستخدم الإحصاءة

$$T = \frac{1 - \frac{s^2}{\bar{x}}}{\sqrt{\frac{2}{n-1}}} \quad (١٢)$$

التي يمكن إثبات أن توزيعها يطابق توزيع  $T$  بدرجات حرية عددها  $n - 1$  . ويجمل بنا أن نستخدم اختباراً ذا جانب واحد لأن المطلوب هنا معرفة ما إذا كانت



ح أكبر من الواحد ( أو أقل من الواحد ) للمساهمة في تفسير ما قد يوجد من أنماط .

مثال ( ٦ - ٥ ) :

اعتبر عينة عشوائية حجمها ٤٠٠ من الأحداث النادرة . اختبر عشوائية ( استقلال ) هذه الأحداث عند مستوى الدلالة ٠,٠٥ في كل من الحالتين الآتيتين :

( أ ) إذا كان الوسط الحسابي للعينة ١,٨ والتباين ١,٩٦٥ .

( ب ) إذا كان الوسط الحسابي للعينة ٣,٣ والتباين ٢,١٥ .

الحل :

$$( أ ) \text{ ح } = \frac{1,965}{1,8} = \sqrt{1,092}$$

الفرض الصفري ف :  $1 = \text{ح}$

الفرض الآخر ف :  $1 < \text{ح}$

من (١٢) نجد أن :

$$ت = \frac{1,092 - 1}{\frac{1}{399} \sqrt{\frac{1}{399}}} = \frac{0,092}{0,071} = 1,296$$

من الجدول (٧) نجد أن ت  $1,640 - 1,296 = 0,344$

بما أن  $1,296 > 1,640$  لا يكون لدينا دليل ضد الفرض الصفري  $1 = \text{ح}$  ولا يسعنا إلا قبوله . وهذا يعني أن الأحداث تتوزع بواسونياً وبالتالي تقع عشوائياً مستقلة عن بعضها البعض .

$$( ب ) \text{ ح } = \frac{2,15}{3,3} = \sqrt{0,652}$$



اختبار الفرض المذكور يمكن أن يؤسس على المقارنة بين هذين النوعين من التكرارات .

والإحصاء المناسبة لذلك هي :

$$\chi^2 = \sum_{i=1}^m \frac{(n_i - n_i')^2}{n_i'} \quad \text{حيث } n_i' = n_i \quad (13)$$

$$\text{وحيث } n_i' = n_i = n \quad (14)$$

التي يمكن إثبات أن توزيعها يقترب من توزيع  $\chi^2$  بـ درجات حرية  $(m - 1)$  كلما اقتربت  $n$  من اللانهاية .

ولصحة استخدام هذه الإحصاء ينبغي أن تتحقق الشروط الآتية :

(أ) أن تكون العينة عشوائية .

(ب) ألا يقل حجم العينة عن ٤٠ (لأن توزيع الإحصاء  $\chi^2$  تقريبي) .

(ج) ألا تقل أى  $n_i$  عن خمسة ، على أنه إذا وجدت قيمة  $n_i$  تقل عن خمسة تضم هذه القيمة إلى قيمة مجاورة لها ، ونفعل ذلك لقيم  $n_i$  المناظرة .

وجدير بالملاحظة أن (١٣) لا تساوى الصفر إلا إذا كانت  $n_i' = n_i$  لجميع

$i$  وتزيد قيمتها كلما كبرت الفروق بين  $n_i'$  ،  $n_i$  .

إن الفرض الصفري  $H_0$  هنا هو عدم وجود فرق بين التكرارات المشاهدة

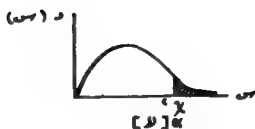
$n_i$  والتكرارات  $n_i'$  المتوقعة لها . أما الفرض الآخر  $H_1$  فهو أن هنالك تفاوتاً

بين هذين النوعين من التكرارات يجعل القيمة المشاهدة للإحصاء (١٣) كبيرة

كبراً ذا دلالة . وعلى ذلك تأخذ المنطقة المرحجة الصورة الآتية :

$$\{ \chi^2_{\text{ملاحظة}} < \chi^2 : \chi^2 \} = \alpha$$

حيث  $\alpha$  ترمز إلى درجات الحرية .



الشكل (٩-٣)

### حساب درجات الحرية :

نعلم أن توزيع  $\chi^2$  يتوقف كلية على عدد درجات الحرية  $\nu$  . ولايجاد هذا العدد نحسب عدد المقارنات المستقلة بين  $k_r$  ،  $q_r$  ( حيث  $r = 1, 2, \dots, m$  ) ونظراً لأن حجم العينة  $n$  هو مقدار ثابت معروف مقدماً فإن التكرارات  $k_r$  ( أو  $q_r$  ) لا تكون جميعها مستقلة إذ نستطيع إذا عرفنا  $(m - 1)$  منها أن نعرف من المتساوية (١٤) التكرار الباقي . وعلى هذا فإن عدد المقارنات المستقلة هو  $m - 1$  . فإذا كانت القيمة المشاهدة للإحصاء (١٣) قد حسبت اعتماداً على حجم العينة فقط فإن  $\nu = m - 1$  .

وكقاعدة عامة كل قيد خطي تنقيد به القيم الداخلة في تركيب الإحصاء (١٣) ينقص واحداً من درجات الحرية ، ولقد كان ثبات حجم العينة هو أحد هذه القيود . وإذا كنا قد احتجنا لحساب القيم المتوقعة إلى واحد أو أكثر من أدلة المجتمع المجهولة واضطررنا إلى تقديرها من العينة فإننا نكون قد قيدنا أنفسنا بالعينة التي بأيدينا وبالتالي فإن كلا من هذه التقديرات يشكل قيداً على التوزيع ينقص واحداً من درجات الحرية . وكطريقة إجرائية سهلة ، نحسب عدد درجات الحرية كالآتي :

$\nu = \text{عدد المقارنات المستقلة} - \text{عدد الأدلة التي قدرت من العينة}$  (١٥)

فيما يلي ثلاثة تطبيقات تستخدم اختبار  $\chi^2$  بالصيغة المبينة في (١٣) .

(٦ - ٧ - ١) اختبار فرض عن توزيع مجتمع :

مثال (٦ - ٦) :

في أحد تجارب مندل الشهيرة في تهجين النباتات نتج من نبات البسلة ما يلي :  
 ٣١٥ نباتاً مستديراً أصفراً ، ١٠٨ مستديراً أخضر ، ١٠١ مجعداً أصفراً ، ٣٢  
 مجعداً أخضر . وطبقاً لنظرية مندل يجب أن تكون أعداد هذه الأنواع متناسبة  
 مع ٩ : ٣ : ٣ : ١ . هل هناك دليل يدعو إلى الشك في نظرية مندل من واقع  
 هذه البيانات ؟ استخدم مستوى الدلالة ٠,٠٥ .

الحل :

لدينا ٣١٥ + ١٠٨ + ١٠١ + ٣٢ = ٥٥٦ نباتاً ( حجم العينة) .

إذا كانت نظرية مندل صحيحة ، تتوزع هذه النباتات بالنسب ٩ : ٣ : ٣ : ١  
 كالآتي :

$$\text{مستدير أصفر} = \frac{9}{16} \times 556 = 312,75$$

$$\text{مستدير أخضر} = \frac{3}{16} \times 556 = 104,25$$

$$\text{مجعد أصفر} = \frac{3}{16} \times 556 = 104,25$$

$$\text{مجعد أخضر} = \frac{1}{16} \times 556 = 34,75$$

التكرارات المشاهدة لـ : ٣١٥ ١٠٨ ١٠١ ٣٢

التكرارات المتوقعة لـ : ٣١٢,٧٥ ١٠٤,٢٥ ١٠٤,٢٥ ٣٤,٧٥

الفرض الصفري  $H_0$  : نظرية مندل صحيحة .

الفرض الآخر  $H_1$  : نظرية مندل غير صحيحة وبالتالي تكون القيم

المشاهدة للإحصاء (١٣) كبيرة كبراً ذا دلالة .

$$\chi^2 = \frac{(34,75-32)^2}{34,75} + \frac{(104,25-101)^2}{104,25} + \frac{(104,25-108)^2}{104,25} + \frac{(312,75-315)^2}{312,75} = 0,47$$

$$\chi^2 = 1 - \epsilon = 3$$

$$\text{من جدول } \chi^2 \text{ نجد أن } \chi^2_{0.05, 100} = 7.82$$

العدد ٠,٤٧ يقل كثيراً عن ٧,٨٢ ( لا يقع في منطقة الرفض ) وإذن لا يمكن رفض الفرض الصفري عند مستوى الدلالة ٠,٠٥ وبالتالي نقبله ونقرر أنه لا يوجد دليل يدعو إلى الشك في نظرية مندل وأن نتيجة هذه التجربة تدعم هذه النظرية .

مثال (٦ - ٧) :

المفروض في جداول الأعداد العشوائية أن تظهر الأرقام ٠ ، ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ ، ٧ ، ٨ ، ٩ باحتمالات متساوية . بفحص إحدى صفحات أحد هذه الجداول وجد التوزيع الآتي :

الرقم	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
التكرار	١٧	٣١	٢٩	١٨	١٤	٢٠	٣٥	٣٠	٢٠	٣٦

اختبر الفرض أن الأرقام العشرة تظهر باحتمالات متساوية مستخدماً مستوى الدلالة ٠,٠١

الحل :

مجموع الأرقام التي بالصفحة = ١٧ + ٣١ + ٠٠٠ + ٣٦ = ٢٥٠ رقماً  
إذا كانت الأرقام العشرة تظهر باحتمالات متساوية فإن كلا منها يجب أن يظهر ٢٥٠ ÷ ١٠ = ٢٥ مرة .

أى أن التكرارات المتوقعة على أساس صحة هذا الفرض هي ٢٥ لكل من الأعداد العشرة .

$$\chi^2 = \frac{(25-36)^2}{25} + \dots + \frac{(25-31)^2}{25} + \frac{(25-17)^2}{25}$$

$$= 23,3 \quad \text{حيث } 9 = 1 - 0,1$$

$$21,666 = \chi^2_{(9),0,1}$$

بما أن 23,3 أكبر من 21,666 نرفض الفرض الصفري عن تساوى احتمالات ظهور الأرقام العشرة عند مستوى الدلالة 0,01 ويتضمن هذا أن هناك شك في دقة هذا الجدول .

(٦ - ٧ - ٢) اختبار حسن المطابقة

#### TEST OF GOODNESS OF FIT

يقصد بهذا الاختبار تحديد مدى ملائمة توزيع نظري معروف مثل ذى الحدين وبواسون والمتعدل ، لتوزيع تكرارى مشاهد في عينة ، بهدف التحقق من صلاحيته كتوزيع للمجتمع الذى أخذت منه العينة .

مثال (٦ - ٨) :

اختبر الفرض القائل أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال من واقع البيانات المعطاة في المثال (٣ - ٥) عن العائلات ذوات الأربعة الأطفال وهى :

عدد الذكور في العائلة	٥	٤	٣	٢	١	٠	س :
عدد العائلات	١٣	٤٣	٩٢	١١٢	٦٠	ل :	

الحل :

على فرض صحة الفرض الصفري أن إنجاب الذكور وإنجاب الإناث متساويا الاحتمال يكون احتمال إنجاب الذكور  $H = \frac{1}{2}$  ويكون توزيع عدد الذكور في العائلات

ذوات الأربعة الأظفار هو توزيع ذى الحدين دليله ٤ ، ١-٣ . وقد وجدنا في المثال (٣ - ٥) أن هذا التوزيع هو :

عدد الذكور في العائلة سر : ٠ ١ ٢ ٣ ٤  
عدد العائلات سر : ٢٠ ٨٠ ١٢٠ ٨٠ ٢٠

$$\chi^2 = \frac{(20-60)^2}{20} + \frac{(80-112)^2}{80} + \frac{(120-92)^2}{120} + \frac{(80-42)^2}{80} + \frac{(20-12)^2}{20}$$

$$= 118,89 \quad \text{حيث } 1 - 0 = 1 \quad 4 =$$

$$\chi^2_{(4),0.01} = 13,277$$

بما أن  $118,89 < 13,277$  نرفض صحة الفرض أن احتمال إنجاب الذكور يساوى احتمال إنجاب الإناث ، وذلك عند مستوى الدلالة ٠,٠١ .

**مثال (٦ - ٩) اختبار الاعتدالية : TEST OF NORMALITY**

اختبر ما إذا كان من الممكن اعتبار العينة المعطاة في المثال (٤ - ٤) مأخوذة من مجتمع معتدل .

**الحل :**

في ذلك المثال وقفنا توزيعاً معتدلاً للتوزيع المعطى ووجدنا ما يلي :

التكرارات المشاهدة لـ	٦	٦	١١	١٤	١٦	١٥	٨
التكرارات المتوقعة لـ	٦,٨١	٦,٥٤	٩,٦١	١٢,٩٨	١٣,٦٦	١٥,٥٧	١٢,٤٧
التكرارات المشاهدة لـ	١٠	٨	٦				
التكرارات المتوقعة لـ	٩,٤٤	٦,٣٧	٦,٥٥				



$$\chi^2 = \frac{(1,00-1)^2}{1,00} + \frac{(1,27-1)^2}{1,27} + \dots + \frac{(1,04-1)^2}{1,04} + \frac{(1,81-1)^2}{1,81} = 2,9427$$

بما أننا كنا قد قدرنا من العينة اثنتين من أدلة المجتمع هما الوسط الحسابي والانحراف المعياري فإن عدد درجات الحرية  $\nu = (10 - 1) = 9$ .

$$\chi^2_{0.05, 9} = 14,067$$

بما أن  $2,9427 < 14,067$  لا نستطيع رفض الفرض الصفري عن اعتدالية المجتمع عند مستوى الدلالة 0,05.

### ملاحظة (١) :

إن عيب اختبار  $\chi^2$  لحسن المطابقة أنه لا يهتم بإشارات الفروق بين التكرارات المشاهدة  $h_i$  والتكرارات النظرية  $e_i$  (لأنه يأخذ مربعات هذه الفروق) ففي توفيق توزيع معتدل مثلاً قد يحدث أن تكون قيم  $h_i$  أصغر من قيم  $e_i$  في الجزء الأوسط من التوزيع وتكون أكبر منها في الطرفين وهذا يعني أن التوزيع المشاهد أكثر انسياطاً من التوزيع المعتدل أى يأخذ شكلاً يختلف عن شكل التوزيع المعتدل. ومع هذا قد تكون قيمة  $\chi^2$  ليست بذات دلالة وتدعونا إلى اعتبار أن المجتمع معتدل. ولهذا لا يكون الاختبار مناسباً إلا إذا لم يكن هناك نمط معين للفروق بين  $h_i$  و  $e_i$ .

### (٦ - ٧ - ٣) اختبار استقلال خاصيتين :

#### TEST OF INDEPENDENCE

في بعض الدراسات تتناول مجتمعاً صنفت عناصره بحسب خاصية ما إلى م من الأقسام المنفصلة ١، ٢، ٣، ٤، ٥، ٦، ٧، ٨، ٩، ١٠، ١١، ١٢، ١٣، ١٤، ١٥، ١٦، ١٧، ١٨، ١٩، ٢٠، ٢١، ٢٢، ٢٣، ٢٤، ٢٥، ٢٦، ٢٧، ٢٨، ٢٩، ٣٠، ٣١، ٣٢، ٣٣، ٣٤، ٣٥، ٣٦، ٣٧، ٣٨، ٣٩، ٤٠، ٤١، ٤٢، ٤٣، ٤٤، ٤٥، ٤٦، ٤٧، ٤٨، ٤٩، ٥٠، ٥١، ٥٢، ٥٣، ٥٤، ٥٥، ٥٦، ٥٧، ٥٨، ٥٩، ٦٠، ٦١، ٦٢، ٦٣، ٦٤، ٦٥، ٦٦، ٦٧، ٦٨، ٦٩، ٧٠، ٧١، ٧٢، ٧٣، ٧٤، ٧٥، ٧٦، ٧٧، ٧٨، ٧٩، ٨٠، ٨١، ٨٢، ٨٣، ٨٤، ٨٥، ٨٦، ٨٧، ٨٨، ٨٩، ٩٠، ٩١، ٩٢، ٩٣، ٩٤، ٩٥، ٩٦، ٩٧، ٩٨، ٩٩، ١٠٠، ١٠١، ١٠٢، ١٠٣، ١٠٤، ١٠٥، ١٠٦، ١٠٧، ١٠٨، ١٠٩، ١١٠، ١١١، ١١٢، ١١٣، ١١٤، ١١٥، ١١٦، ١١٧، ١١٨، ١١٩، ١٢٠، ١٢١، ١٢٢، ١٢٣، ١٢٤، ١٢٥، ١٢٦، ١٢٧، ١٢٨، ١٢٩، ١٣٠، ١٣١، ١٣٢، ١٣٣، ١٣٤، ١٣٥، ١٣٦، ١٣٧، ١٣٨، ١٣٩، ١٤٠، ١٤١، ١٤٢، ١٤٣، ١٤٤، ١٤٥، ١٤٦، ١٤٧، ١٤٨، ١٤٩، ١٥٠، ١٥١، ١٥٢، ١٥٣، ١٥٤، ١٥٥، ١٥٦، ١٥٧، ١٥٨، ١٥٩، ١٦٠، ١٦١، ١٦٢، ١٦٣، ١٦٤، ١٦٥، ١٦٦، ١٦٧، ١٦٨، ١٦٩، ١٧٠، ١٧١، ١٧٢، ١٧٣، ١٧٤، ١٧٥، ١٧٦، ١٧٧، ١٧٨، ١٧٩، ١٨٠، ١٨١، ١٨٢، ١٨٣، ١٨٤، ١٨٥، ١٨٦، ١٨٧، ١٨٨، ١٨٩، ١٩٠، ١٩١، ١٩٢، ١٩٣، ١٩٤، ١٩٥، ١٩٦، ١٩٧، ١٩٨، ١٩٩، ٢٠٠، ٢٠١، ٢٠٢، ٢٠٣، ٢٠٤، ٢٠٥، ٢٠٦، ٢٠٧، ٢٠٨، ٢٠٩، ٢١٠، ٢١١، ٢١٢، ٢١٣، ٢١٤، ٢١٥، ٢١٦، ٢١٧، ٢١٨، ٢١٩، ٢٢٠، ٢٢١، ٢٢٢، ٢٢٣، ٢٢٤، ٢٢٥، ٢٢٦، ٢٢٧، ٢٢٨، ٢٢٩، ٢٣٠، ٢٣١، ٢٣٢، ٢٣٣، ٢٣٤، ٢٣٥، ٢٣٦، ٢٣٧، ٢٣٨، ٢٣٩، ٢٤٠، ٢٤١، ٢٤٢، ٢٤٣، ٢٤٤، ٢٤٥، ٢٤٦، ٢٤٧، ٢٤٨، ٢٤٩، ٢٥٠، ٢٥١، ٢٥٢، ٢٥٣، ٢٥٤، ٢٥٥، ٢٥٦، ٢٥٧، ٢٥٨، ٢٥٩، ٢٦٠، ٢٦١، ٢٦٢، ٢٦٣، ٢٦٤، ٢٦٥، ٢٦٦، ٢٦٧، ٢٦٨، ٢٦٩، ٢٧٠، ٢٧١، ٢٧٢، ٢٧٣، ٢٧٤، ٢٧٥، ٢٧٦، ٢٧٧، ٢٧٨، ٢٧٩، ٢٨٠، ٢٨١، ٢٨٢، ٢٨٣، ٢٨٤، ٢٨٥، ٢٨٦، ٢٨٧، ٢٨٨، ٢٨٩، ٢٩٠، ٢٩١، ٢٩٢، ٢٩٣، ٢٩٤، ٢٩٥، ٢٩٦، ٢٩٧، ٢٩٨، ٢٩٩، ٣٠٠، ٣٠١، ٣٠٢، ٣٠٣، ٣٠٤، ٣٠٥، ٣٠٦، ٣٠٧، ٣٠٨، ٣٠٩، ٣١٠، ٣١١، ٣١٢، ٣١٣، ٣١٤، ٣١٥، ٣١٦، ٣١٧، ٣١٨، ٣١٩، ٣٢٠، ٣٢١، ٣٢٢، ٣٢٣، ٣٢٤، ٣٢٥، ٣٢٦، ٣٢٧، ٣٢٨، ٣٢٩، ٣٣٠، ٣٣١، ٣٣٢، ٣٣٣، ٣٣٤، ٣٣٥، ٣٣٦، ٣٣٧، ٣٣٨، ٣٣٩، ٣٤٠، ٣٤١، ٣٤٢، ٣٤٣، ٣٤٤، ٣٤٥، ٣٤٦، ٣٤٧، ٣٤٨، ٣٤٩، ٣٥٠، ٣٥١، ٣٥٢، ٣٥٣، ٣٥٤، ٣٥٥، ٣٥٦، ٣٥٧، ٣٥٨، ٣٥٩، ٣٦٠، ٣٦١، ٣٦٢، ٣٦٣، ٣٦٤، ٣٦٥، ٣٦٦، ٣٦٧، ٣٦٨، ٣٦٩، ٣٧٠، ٣٧١، ٣٧٢، ٣٧٣، ٣٧٤، ٣٧٥، ٣٧٦، ٣٧٧، ٣٧٨، ٣٧٩، ٣٨٠، ٣٨١، ٣٨٢، ٣٨٣، ٣٨٤، ٣٨٥، ٣٨٦، ٣٨٧، ٣٨٨، ٣٨٩، ٣٩٠، ٣٩١، ٣٩٢، ٣٩٣، ٣٩٤، ٣٩٥، ٣٩٦، ٣٩٧، ٣٩٨، ٣٩٩، ٤٠٠، ٤٠١، ٤٠٢، ٤٠٣، ٤٠٤، ٤٠٥، ٤٠٦، ٤٠٧، ٤٠٨، ٤٠٩، ٤١٠، ٤١١، ٤١٢، ٤١٣، ٤١٤، ٤١٥، ٤١٦، ٤١٧، ٤١٨، ٤١٩، ٤٢٠، ٤٢١، ٤٢٢، ٤٢٣، ٤٢٤، ٤٢٥، ٤٢٦، ٤٢٧، ٤٢٨، ٤٢٩، ٤٣٠، ٤٣١، ٤٣٢، ٤٣٣، ٤٣٤، ٤٣٥، ٤٣٦، ٤٣٧، ٤٣٨، ٤٣٩، ٤٤٠، ٤٤١، ٤٤٢، ٤٤٣، ٤٤٤، ٤٤٥، ٤٤٦، ٤٤٧، ٤٤٨، ٤٤٩، ٤٥٠، ٤٥١، ٤٥٢، ٤٥٣، ٤٥٤، ٤٥٥، ٤٥٦، ٤٥٧، ٤٥٨، ٤٥٩، ٤٦٠، ٤٦١، ٤٦٢، ٤٦٣، ٤٦٤، ٤٦٥، ٤٦٦، ٤٦٧، ٤٦٨، ٤٦٩، ٤٧٠، ٤٧١، ٤٧٢، ٤٧٣، ٤٧٤، ٤٧٥، ٤٧٦، ٤٧٧، ٤٧٨، ٤٧٩، ٤٨٠، ٤٨١، ٤٨٢، ٤٨٣، ٤٨٤، ٤٨٥، ٤٨٦، ٤٨٧، ٤٨٨، ٤٨٩، ٤٩٠، ٤٩١، ٤٩٢، ٤٩٣، ٤٩٤، ٤٩٥، ٤٩٦، ٤٩٧، ٤٩٨، ٤٩٩، ٥٠٠، ٥٠١، ٥٠٢، ٥٠٣، ٥٠٤، ٥٠٥، ٥٠٦، ٥٠٧، ٥٠٨، ٥٠٩، ٥١٠، ٥١١، ٥١٢، ٥١٣، ٥١٤، ٥١٥، ٥١٦، ٥١٧، ٥١٨، ٥١٩، ٥٢٠، ٥٢١، ٥٢٢، ٥٢٣، ٥٢٤، ٥٢٥، ٥٢٦، ٥٢٧، ٥٢٨، ٥٢٩، ٥٣٠، ٥٣١، ٥٣٢، ٥٣٣، ٥٣٤، ٥٣٥، ٥٣٦، ٥٣٧، ٥٣٨، ٥٣٩، ٥٤٠، ٥٤١، ٥٤٢، ٥٤٣، ٥٤٤، ٥٤٥، ٥٤٦، ٥٤٧، ٥٤٨، ٥٤٩، ٥٥٠، ٥٥١، ٥٥٢، ٥٥٣، ٥٥٤، ٥٥٥، ٥٥٦، ٥٥٧، ٥٥٨، ٥٥٩، ٥٦٠، ٥٦١، ٥٦٢، ٥٦٣، ٥٦٤، ٥٦٥، ٥٦٦، ٥٦٧، ٥٦٨، ٥٦٩، ٥٧٠، ٥٧١، ٥٧٢، ٥٧٣، ٥٧٤، ٥٧٥، ٥٧٦، ٥٧٧، ٥٧٨، ٥٧٩، ٥٨٠، ٥٨١، ٥٨٢، ٥٨٣، ٥٨٤، ٥٨٥، ٥٨٦، ٥٨٧، ٥٨٨، ٥٨٩، ٥٩٠، ٥٩١، ٥٩٢، ٥٩٣، ٥٩٤، ٥٩٥، ٥٩٦، ٥٩٧، ٥٩٨، ٥٩٩، ٦٠٠، ٦٠١، ٦٠٢، ٦٠٣، ٦٠٤، ٦٠٥، ٦٠٦، ٦٠٧، ٦٠٨، ٦٠٩، ٦١٠، ٦١١، ٦١٢، ٦١٣، ٦١٤، ٦١٥، ٦١٦، ٦١٧، ٦١٨، ٦١٩، ٦٢٠، ٦٢١، ٦٢٢، ٦٢٣، ٦٢٤، ٦٢٥، ٦٢٦، ٦٢٧، ٦٢٨، ٦٢٩، ٦٣٠، ٦٣١، ٦٣٢، ٦٣٣، ٦٣٤، ٦٣٥، ٦٣٦، ٦٣٧، ٦٣٨، ٦٣٩، ٦٤٠، ٦٤١، ٦٤٢، ٦٤٣، ٦٤٤، ٦٤٥، ٦٤٦، ٦٤٧، ٦٤٨، ٦٤٩، ٦٥٠، ٦٥١، ٦٥٢، ٦٥٣، ٦٥٤، ٦٥٥، ٦٥٦، ٦٥٧، ٦٥٨، ٦٥٩، ٦٦٠، ٦٦١، ٦٦٢، ٦٦٣، ٦٦٤، ٦٦٥، ٦٦٦، ٦٦٧، ٦٦٨، ٦٦٩، ٦٧٠، ٦٧١، ٦٧٢، ٦٧٣، ٦٧٤، ٦٧٥، ٦٧٦، ٦٧٧، ٦٧٨، ٦٧٩، ٦٨٠، ٦٨١، ٦٨٢، ٦٨٣، ٦٨٤، ٦٨٥، ٦٨٦، ٦٨٧، ٦٨٨، ٦٨٩، ٦٩٠، ٦٩١، ٦٩٢، ٦٩٣، ٦٩٤، ٦٩٥، ٦٩٦، ٦٩٧، ٦٩٨، ٦٩٩، ٧٠٠، ٧٠١، ٧٠٢، ٧٠٣، ٧٠٤، ٧٠٥، ٧٠٦، ٧٠٧، ٧٠٨، ٧٠٩، ٧١٠، ٧١١، ٧١٢، ٧١٣، ٧١٤، ٧١٥، ٧١٦، ٧١٧، ٧١٨، ٧١٩، ٧٢٠، ٧٢١، ٧٢٢، ٧٢٣، ٧٢٤، ٧٢٥، ٧٢٦، ٧٢٧، ٧٢٨، ٧٢٩، ٧٣٠، ٧٣١، ٧٣٢، ٧٣٣، ٧٣٤، ٧٣٥، ٧٣٦، ٧٣٧، ٧٣٨، ٧٣٩، ٧٤٠، ٧٤١، ٧٤٢، ٧٤٣، ٧٤٤، ٧٤٥، ٧٤٦، ٧٤٧، ٧٤٨، ٧٤٩، ٧٥٠، ٧٥١، ٧٥٢، ٧٥٣، ٧٥٤، ٧٥٥، ٧٥٦، ٧٥٧، ٧٥٨، ٧٥٩، ٧٦٠، ٧٦١، ٧٦٢، ٧٦٣، ٧٦٤، ٧٦٥، ٧٦٦، ٧٦٧، ٧٦٨، ٧٦٩، ٧٧٠، ٧٧١، ٧٧٢، ٧٧٣، ٧٧٤، ٧٧٥، ٧٧٦، ٧٧٧، ٧٧٨، ٧٧٩، ٧٨٠، ٧٨١، ٧٨٢، ٧٨٣، ٧٨٤، ٧٨٥، ٧٨٦، ٧٨٧، ٧٨٨، ٧٨٩، ٧٩٠، ٧٩١، ٧٩٢، ٧٩٣، ٧٩٤، ٧٩٥، ٧٩٦، ٧٩٧، ٧٩٨، ٧٩٩، ٨٠٠، ٨٠١، ٨٠٢، ٨٠٣، ٨٠٤، ٨٠٥، ٨٠٦، ٨٠٧، ٨٠٨، ٨٠٩، ٨١٠، ٨١١، ٨١٢، ٨١٣، ٨١٤، ٨١٥، ٨١٦، ٨١٧، ٨١٨، ٨١٩، ٨٢٠، ٨٢١، ٨٢٢، ٨٢٣، ٨٢٤، ٨٢٥، ٨٢٦، ٨٢٧، ٨٢٨، ٨٢٩، ٨٣٠، ٨٣١، ٨٣٢، ٨٣٣، ٨٣٤، ٨٣٥، ٨٣٦، ٨٣٧، ٨٣٨، ٨٣٩، ٨٤٠، ٨٤١، ٨٤٢، ٨٤٣، ٨٤٤، ٨٤٥، ٨٤٦، ٨٤٧، ٨٤٨، ٨٤٩، ٨٥٠، ٨٥١، ٨٥٢، ٨٥٣، ٨٥٤، ٨٥٥، ٨٥٦، ٨٥٧، ٨٥٨، ٨٥٩، ٨٦٠، ٨٦١، ٨٦٢، ٨٦٣، ٨٦٤، ٨٦٥، ٨٦٦، ٨٦٧، ٨٦٨، ٨٦٩، ٨٧٠، ٨٧١، ٨٧٢، ٨٧٣، ٨٧٤، ٨٧٥، ٨٧٦، ٨٧٧، ٨٧٨، ٨٧٩، ٨٨٠، ٨٨١، ٨٨٢، ٨٨٣، ٨٨٤، ٨٨٥، ٨٨٦، ٨٨٧، ٨٨٨، ٨٨٩، ٨٩٠، ٨٩١، ٨٩٢، ٨٩٣، ٨٩٤، ٨٩٥، ٨٩٦، ٨٩٧، ٨٩٨، ٨٩٩، ٩٠٠، ٩٠١، ٩٠٢، ٩٠٣، ٩٠٤، ٩٠٥، ٩٠٦، ٩٠٧، ٩٠٨، ٩٠٩، ٩١٠، ٩١١، ٩١٢، ٩١٣، ٩١٤، ٩١٥، ٩١٦، ٩١٧، ٩١٨، ٩١٩، ٩٢٠، ٩٢١، ٩٢٢، ٩٢٣، ٩٢٤، ٩٢٥، ٩٢٦، ٩٢٧، ٩٢٨، ٩٢٩، ٩٣٠، ٩٣١، ٩٣٢، ٩٣٣، ٩٣٤، ٩٣٥، ٩٣٦، ٩٣٧، ٩٣٨، ٩٣٩، ٩٤٠، ٩٤١، ٩٤٢، ٩٤٣، ٩٤٤، ٩٤٥، ٩٤٦، ٩٤٧، ٩٤٨، ٩٤٩، ٩٥٠، ٩٥١، ٩٥٢، ٩٥٣، ٩٥٤، ٩٥٥، ٩٥٦، ٩٥٧، ٩٥٨، ٩٥٩، ٩٦٠، ٩٦١، ٩٦٢، ٩٦٣، ٩٦٤، ٩٦٥، ٩٦٦، ٩٦٧، ٩٦٨، ٩٦٩، ٩٧٠، ٩٧١، ٩٧٢، ٩٧٣، ٩٧٤، ٩٧٥، ٩٧٦، ٩٧٧، ٩٧٨، ٩٧٩، ٩٨٠، ٩٨١، ٩٨٢، ٩٨٣، ٩٨٤، ٩٨٥، ٩٨٦، ٩٨٧، ٩٨٨، ٩٨٩، ٩٩٠، ٩٩١، ٩٩٢، ٩٩٣، ٩٩٤، ٩٩٥، ٩٩٦، ٩٩٧، ٩٩٨، ٩٩٩، ١٠٠٠، ١٠٠١، ١٠٠٢، ١٠٠٣، ١٠٠٤، ١٠٠٥، ١٠٠٦، ١٠٠٧، ١٠٠٨، ١٠٠٩، ١٠١٠، ١٠١١، ١٠١٢، ١٠١٣، ١٠١٤، ١٠١٥، ١٠١٦، ١٠١٧، ١٠١٨، ١٠١٩، ١٠٢٠، ١٠٢١، ١٠٢٢، ١٠٢٣، ١٠٢٤، ١٠٢٥، ١٠٢٦، ١٠٢٧، ١٠٢٨، ١٠٢٩، ١٠٣٠، ١٠٣١، ١٠٣٢، ١٠٣٣، ١٠٣٤، ١٠٣٥، ١٠٣٦، ١٠٣٧، ١٠٣٨، ١٠٣٩، ١٠٤٠، ١٠٤١، ١٠٤٢، ١٠٤٣، ١٠٤٤، ١٠٤٥، ١٠٤٦، ١٠٤٧، ١٠٤٨، ١٠٤٩، ١٠٥٠، ١٠٥١، ١٠٥٢، ١٠٥٣، ١٠٥٤، ١٠٥٥، ١٠٥٦، ١٠٥٧، ١٠٥٨، ١٠٥٩، ١٠٦٠، ١٠٦١، ١٠٦٢، ١٠٦٣، ١٠٦٤، ١٠٦٥، ١٠٦٦، ١٠٦٧، ١٠٦٨، ١٠٦٩، ١٠٧٠، ١٠٧١، ١٠٧٢، ١٠٧٣، ١٠٧٤، ١٠٧٥، ١٠٧٦، ١٠٧٧، ١٠٧٨، ١٠٧٩، ١٠٨٠، ١٠٨١، ١٠٨٢، ١٠٨٣، ١٠٨٤، ١٠٨٥، ١٠٨٦، ١٠٨٧، ١٠٨٨، ١٠٨٩، ١٠٩٠، ١٠٩١، ١٠٩٢، ١٠٩٣، ١٠٩٤، ١٠٩٥، ١٠٩٦، ١٠٩٧، ١٠٩٨، ١٠٩٩، ١١٠٠، ١١٠١، ١١٠٢، ١١٠٣، ١١٠٤، ١١٠٥، ١١٠٦، ١١٠٧، ١١٠٨، ١١٠٩، ١١١٠، ١١١١، ١١١٢، ١١١٣، ١١١٤، ١١١٥، ١١١٦، ١١١٧، ١١١٨، ١١١٩، ١١٢٠، ١١٢١، ١١٢٢، ١١٢٣، ١١٢٤، ١١٢٥، ١١٢٦، ١١٢٧، ١١٢٨، ١١٢٩، ١١٣٠، ١١٣١، ١١٣٢، ١١٣٣، ١١٣٤، ١١٣٥، ١١٣٦، ١١٣٧، ١١٣٨، ١١٣٩، ١١٤٠، ١١٤١، ١١٤٢، ١١٤٣، ١١٤٤، ١١٤٥، ١١٤٦، ١١٤٧، ١١٤٨، ١١٤٩، ١١٥٠، ١١٥١، ١١٥٢، ١١٥٣، ١١٥٤، ١١٥٥، ١١٥٦، ١١٥٧، ١١٥٨، ١١٥٩، ١١٦٠، ١١٦١، ١١٦٢، ١١٦٣، ١١٦٤، ١١٦٥، ١١٦٦، ١١٦٧، ١١٦٨، ١١٦٩، ١١٧٠، ١١٧١، ١١٧٢، ١١٧٣، ١١٧٤، ١١٧٥، ١١٧٦، ١١٧٧، ١١٧٨، ١١٧٩، ١١٨٠، ١١٨١، ١١٨٢، ١١٨٣، ١١٨٤، ١١٨٥، ١١٨٦، ١١٨٧، ١١٨٨، ١١٨٩، ١١٩٠، ١١٩١، ١١٩٢، ١١٩٣، ١١٩٤، ١١٩٥، ١١٩٦، ١١٩٧، ١١٩٨، ١١٩٩، ١٢٠٠، ١٢٠١، ١٢٠٢، ١٢٠٣، ١٢٠٤، ١٢٠٥، ١٢٠٦، ١٢٠٧، ١٢٠٨، ١٢٠٩، ١٢١٠، ١٢١١، ١٢١٢، ١٢١٣، ١٢١٤، ١٢١٥، ١٢١٦، ١٢١٧، ١٢١٨، ١٢١٩، ١٢٢٠، ١٢٢١، ١٢٢٢، ١٢٢٣، ١٢٢٤، ١٢٢٥، ١٢٢٦، ١٢٢٧، ١٢٢٨، ١٢٢٩، ١٢٣٠، ١٢٣١، ١٢٣٢، ١٢٣٣، ١٢٣٤، ١٢٣٥، ١٢٣٦، ١٢٣٧، ١٢٣٨، ١٢٣٩، ١٢٤٠، ١٢٤١، ١٢٤٢، ١٢٤٣، ١٢٤٤، ١٢٤٥، ١٢٤٦، ١٢٤٧، ١٢٤٨، ١٢٤٩، ١٢٥٠، ١٢٥١، ١٢٥٢، ١٢٥٣، ١٢٥٤، ١٢٥٥، ١٢٥٦، ١٢٥٧، ١٢٥٨، ١٢٥٩، ١٢٦٠، ١٢٦١، ١٢٦٢، ١٢٦٣، ١٢٦٤، ١٢٦٥، ١٢٦٦، ١٢٦٧، ١٢٦٨، ١٢٦٩، ١٢٧٠، ١٢٧١، ١٢٧٢، ١٢٧٣، ١٢٧٤، ١٢٧٥، ١٢٧٦، ١٢٧٧، ١٢٧٨، ١٢٧٩، ١٢٨٠، ١٢٨١، ١٢٨٢، ١٢٨٣، ١٢٨٤، ١٢٨٥، ١٢٨٦، ١٢٨٧، ١٢٨٨، ١٢٨٩، ١٢٩٠، ١٢٩١، ١٢٩٢، ١٢٩٣، ١٢٩٤، ١٢٩٥، ١٢٩٦، ١٢٩٧، ١٢٩٨، ١٢٩٩، ١٣٠٠، ١٣٠١، ١٣٠٢، ١٣٠٣، ١٣٠٤، ١٣٠٥، ١٣٠٦، ١٣٠٧، ١٣٠٨، ١٣٠٩، ١٣١٠، ١٣١١، ١٣١٢، ١٣١٣، ١٣١٤، ١٣١٥، ١٣١٦، ١٣١٧، ١٣١٨، ١٣١٩، ١٣٢٠، ١٣٢١، ١٣٢٢، ١٣٢٣، ١٣٢٤، ١٣٢٥، ١٣٢٦، ١٣٢٧، ١٣٢٨، ١٣٢٩، ١٣٣٠، ١٣٣١، ١٣٣٢، ١٣٣٣، ١٣٣٤، ١٣٣٥، ١٣٣٦، ١٣٣٧، ١٣٣٨، ١٣٣٩، ١٣٤٠، ١٣٤١، ١٣٤٢، ١٣٤٣، ١٣٤٤، ١٣٤٥، ١٣٤٦، ١٣٤٧، ١٣٤٨، ١٣٤٩، ١٣٥٠، ١٣٥١، ١٣٥٢، ١٣٥٣، ١٣٥٤، ١٣٥٥، ١٣٥٦، ١٣٥٧، ١٣٥٨، ١٣٥٩، ١٣٦٠، ١٣٦١، ١٣٦٢، ١٣٦٣، ١٣٦٤، ١٣٦٥، ١٣٦٦، ١٣٦٧، ١٣٦٨، ١٣٦٩، ١٣٧٠، ١٣٧١، ١٣٧٢، ١٣٧٣، ١٣٧٤، ١٣٧٥، ١٣٧٦، ١٣٧٧، ١٣٧٨، ١٣٧٩، ١٣٨

في هذه الحال تؤخذ عينة عشوائية وتدرس من حيث عدد العناصر التي تظهر في أقسام الخاصتين وتوضع التكرارات الناتجة بحيث تقترن التكرارات في أقسام الخاصة الأولى بالتكرارات في أقسام الخاصة الثانية فيما يسمى بجدول الاقتران contingency table وهو عبارة عن مصفوفة ذات م صفا تمثل أقسام الخاصة الأولى ، ه عموداً تمثل أقسام الخاصة الثانية ويوصف الجدول حيثذ بأنه جدول اقتران  $2 \times 2$  ه . نضع الفرض الصفري أن الخاصتين مستقلتان ، وعلى أساس صحة هذا الفرض نحسب التكرارات النظرية المناظرة للتكرارات المشاهدة ونضعها في جدول اقتران م  $2 \times 2$  ه أيضا . بمقارنة التكرارات المشاهدة بالتكرارات النظرية نستطيع الحكم على استقلال الخاصتين بواسطة الإحصاءة المعرفة في (١٣) .

#### مثال (٦ - ١٠) :

في تجربة عن الرجال في الأعمار بين ٦٠ ، ٦٤ عاماً صنف الرجال من حيث خاصتي التدخين والوفاة ، وقسمت خاصة التدخين إلى قسمين هما « يدخن ولا يدخن » وقسمت خاصة الوفاة إلى قسمين : رجال لا يزالون على قيد الحياة ورجال توفوا في بحر ٦ سنوات من بدء التجربة . في عينة من ١٤٦٩ رجلا جمعت البيانات في جدول الاقتران  $2 \times 2$  الآتي :

المجموع	عادة التدخين		الوفاة
	لا يدخن	يدخن	
١٧١	١١٧	٥٤	توفي
١٢٩٨	٩٥٠	٣٤٨	حي
١٤٦٩	١٠٦٧	٤٠٢	المجموع

هل هذه البيانات تشير إلى أن الوفاة مستقلة عن عادة التدخين ؟

## الحل :

ف : الوفاة وعادة التدخين خاصتان مستقلتان ، أو لا توجد علاقة بين الخاصتين .

إذا كان هذا الفرض صحيحاً تكون نسبة المتوفين إلى الأحياء واحدة في المجتمع سواء للمدخنين أو لغير المدخنين . ونظراً لأن هذه النسبة غير معروفة نضطر إلى تقديرها من العينة ، ومن المعقول أن نأخذ النسبة التي ظهرت في العينة بين العدد الكلي للمتوفين والعدد الكلي للأحياء وهي ١٧١ : ١٢٩٨ . وباستخدام هذه النسبة نحصل على التكرارات النظرية في الخلايا الأربع وذلك بتقسيم كل من عدد الذين يدخنون وهو ٤٠٢ وعدد الذين لا يدخنون وهو ١٠٦٧ بهذه النسبة فمثلاً :  
التكرار النظري لعدد المتوفين من المدخنين  $= \frac{171}{1298} \times 402 = 52,8$  رجلاً

التكرار النظري لعدد الأحياء من المدخنين  $= \frac{1298}{1298} \times 402 = 402$  رجلاً

وبالمثل بالنسبة لغير المدخنين . وبذلك نحصل على جدول التكرارات النظرية الآتي :

الجموع	عادة التدخين		الوفاة
	لا يدخن	يدخن	
١٧١	١٢٤,٢	٤٦,٨	توفي
١٢٩٨	٩٤٢,٨	٣٥٥,٢	حي
١٤٦٩	١٠٦٧	٤٠٢	المجموع

$$1,73 = \frac{(942,8-900)^2}{942,8} + \dots + \dots + \frac{(46,8-52)^2}{46,8} = \chi^2$$

نحسب عدد درجات الحرية كآلاقي . لدينا ٤ مقارنات ، وبما أن عدد المدخنين ثابت ( وهو ٤٠٢ ) وكذلك عدد غير المدخنين ( وهو ١٠٦٧ ) فإن درجات الحرية تنقص ٢ ، وبما أننا قدرنا دليل واحد للمجتمع من العينة وهو النسبة ١٧١ : ١٢٩٨ فإن درجات الحرية تنقص واحداً آخر .

$$\text{إذن } \nu = 1 - (2 - 4) = 1$$

$$\chi^2_{0.05, 1} = 3.841$$

بما أن  $3.841 > 1.73$  لا نستطيع رفض الفرض الصفري عند المستوى ٠,٠٥ . يمكننا القول بأنه في حدود مجموعة العمر التي درست ، عادة التدخين والمخات مستقلتان ، أى أن التدخين لا يؤثر في الوفاة في حدود هذه التجربة .

مثال (٦ - ١١) :

جدول الاقتران  $3 \times 2$  الآتي يحمل التكرارات المشاهدة في عينة عشوائية من ٩٩ طفلاً حديث الولادة من حيث طول الطفل وطول محيط رأسه بالاستقيمترات ساعة الولادة . ابحت استقلال طول المولود وطول محيط رأسه .

المجموع	طول المولود			محيط الرأس
	٥٥-٥٣	٥٢-٥٠	٤٩-٤٧	
٧٨	٢	٣٦	٤٠	٣٥-٣٢
٢١	٧	١٤	٠	٣٩-٣٦
٩٩	٩	٥٠	٤٠	المجموع

الحل :

كما في المثال السابق ، لو كان طول المولود ومحيط رأسه مستقلين لتوزعت أعداد الأطفال في كل من الأقسام الثلاثة للأطوال بنفس النسبة . ونقدر هذه النسبة من العينة على أنها ٧٨ : ٢١ أى ينبغي أن نقسم كلا من الأعداد ٤٠ ، ٥٠ ، ٩٠ بالنسبة ٧٨ : ٢١ لنحصل على جدول التكرارات النظرية الآتي :

المجموع	طول المولود			محيط الرأس
	٥٥-٥٣	٥٢-٥٠	٤٩-٤٧	
٧٨	٧,١	٣٩,٤	٣١,٥	٣٥-٣٢
٢١	١,٩	١٠,٦	٨,٥	٣٩-٣٦
٩٩	٩	٥٠	٤٠	المجموع

$$\chi^2 = \frac{(1,9-7)^2}{1,9} + \dots + \frac{(31,5-40)^2}{31,5} = 29,377$$

$$v = 1 - (3-1) = 2$$

$$\chi^2_{(2), 0.05} = 5,991$$

نرفض الفرض الصفري عن استقلال طول المولود وطول محيط رأسه عند مستوى الدلالة ٠,٠٥ ونحكم بوجود علاقة بين هاتين الخاصتين .

## ملاحظة (١)

(أ) إذا كانت هناك علاقة موجبة بين الخاصيتين تظهر تكرارات كبيرة نسبياً في خلايا القطر الرئيسى أى في الخلايا العلوية اليمنى وفي الوسط وفي الخلايا السفلية اليسرى كما في هذا المثال .

(ب) وإذا كانت هناك علاقة سالبة تظهر تكرارات كبيرة نسبياً في الخلايا السفلية اليمنى وفي الوسط وفي الخلايا العلوية اليسرى .

(ج) أما إذا كان هناك علاقة صغيرة أو لا توجد علاقة فإن التكرارات تميل إلى أن تكون متناسبة الكثافة بمعنى أن يكون توزيع التكرارات بنفس النسبة التى تظهر في المجاميع الهامشية .

## ملاحظة (٢) :

هناك قاعدة سهلة لإيجاد عدد درجات الحرية في أى جدول اقتران  $m \times h$  حيث  $m < 1$  ،  $h < 1$  وهى :

$$v = (m - 1)(h - 1) \quad (١٦)$$

وتفسير هذه القاعدة أنه مادامت المجاميع الهامشية للأعمدة ثابتة فإننا في ملأ خانات جدول الاقتران يمكننا استنتاج واحد من الأعداد التى بأى عمود من الـ  $h - 1$  الأعداد الأخرى التى بهذا العمود ، وبالمثل يمكن استنتاج واحد من الأعداد التى بأى صف من الـ  $m - 1$  الأعداد الأخرى التى بهذا الصف . وبهذا نكون أحراراً فقط في ملأ  $(m - 1)(h - 1)$  من خلايا الجدول .

## ملاحظة (٣) :

هناك أيضاً قاعدة سهلة لحساب التكرار النظرى المناظر لتكرار مشاهد في أى خلية ، وذلك بضرب المجموعين الهامشين للصف والعمود الذى تقع فيه الخلية

ثم قسمة الناتج على المجموع الكلي للتكرارات ( حجم العينة ) ، فمثلا في المثال (٦ - ١١) الأحمر :

$$\text{التكرار النظري المناظر للعدد ٣٦ هو } \frac{٥٠ \times ٧٨}{٩٩} = ٣٩,٤$$

$$\text{، التكرار النظري المناظر للعدد ٧ هو } \frac{٩ \times ٢١}{٩٩} = ١,٩$$

(٦ - ٧ - ٤) فترات الثقة لتباين مجتمع معطل :

في البند (٦ - ٦ - ٢) أوجدنا صيغة لفترات الثقة للوسط الحسابي لمجتمع معطل واستخدمنا لذلك توزيع ت . أما فترات الثقة لتباين مجتمع معطل فتحاج لاستخدام توزيع  $\chi^2$  ، إذ يمكن إثبات أنه على أساس عينة عشوائية حجمها  $n$  وتباينها  $\sigma^2$  مأخوذة من مجتمع معطل تباينه  $\sigma^2$  فإن المتباينة :

$$(١٧) \quad \frac{\chi^2_{\alpha} (1 - \nu)}{[1 - \nu] \chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{\chi^2_{\alpha} (1 - \nu)}{[1 - \nu] \chi^2_{\frac{\alpha}{2}}}$$

تشكل فترة ثقة بدرجة  $(1 - \alpha)$  لتباين المجتمع . وليس من الضروري هنا أن تكون العينة كبيرة الحجم كما هو الحال في التطبيقات سابقة الذكر لأن التوزيع الذي بنيت عليه هذه الفترة هو توزيع مضبوط طالما كان المجتمع معطلا .

مثال (٦ - ١٢) :

أخذت عينة عشوائية حجمها ١٢ من مجتمع معطل فوجد أن تباينها ٩,٧٣ . أوجد فترة ثقة بدرجة ٩٥٪ لتباين المجتمع .

الحل :

$$\text{لدينا } \alpha = 0,05 \quad \frac{\alpha}{4} = 0,025 \quad 1 - \frac{\alpha}{4} = 0,975 \quad v = 12$$

$$\text{من جدول } \chi^2 \text{ نجد أن } \frac{\alpha}{4} \chi^2 = [1-v] \chi^2 = 21,9$$

$$3,82 = [1-v] \left( \frac{\alpha}{4} \right) \chi^2$$

بالتعويض في (١٧) نجد أن :

$$\text{الحد الأدنى للفترة} = \frac{9,73 \times 11}{21,9} = 4,88$$

$$\text{الحد الأعلى للفترة} = \frac{9,73 \times 11}{3,82} = 28,018$$

إذن الفترة (٤,٨٨ ، ٢٨,٠١٨) هي فترة ثقة بدرجة ٩٥٪ لتباين المجتمع .

## تمارين (٦ - ٢)

( في كل المسائل الآتية اذكر الشروط اللازم توفرها لصحة الحل . )

(١) حسب نظرية مندل عن تهجين النباتات تكون نسبة نبات البسلة الأخضر إلى نبات البسلة الأصفر ٣ : ١ . اختبر هذه النظرية على ضوء البيانات المشاهدة الآتية :

( أ ) وجد في عينة عشوائية أن هناك ٣٥٥ نباتاً أخضر ، ١٢٣ نباتاً أصفر .  
( ب ) وجد في عينة عشوائية أن هناك ٤٢٨ نباتاً أخضر ، ١٥٢ نباتاً أصفر .

(٢) وضعت خمس مصائد في أماكن مختلفة من غابة ما . في فترة ثلاثة أشهر سجلت أعداد الفيران المصيدة كالآتي :

المصيدة	أ	ب	ح	د	هـ
عدد الفيران	٢٣	٧	٢٥	١٩	٢١



اختبر الفرض أنه لا يوجد فرق بين المصايد الخمس في عدد ما تصطاده من الفيران .

(٣) اختبر الفرض أن عدد المواليد اليومية في مجتمع ما ثابت خلال شهور السنة مستخدماً البيانات الآتية :

الشهر : يناير فبراير مارس إبريل مايو يونيو يوليو أغسطس سبتمبر أكتوبر نوفمبر ديسمبر

عدد المواليد : ٨٠ ٧٨ ٨٦ ٨٢ ٨٣ ٧٨ ٧٩ ٧٦ ٧٨ ٧٦ ٧٢ ٧٦

(٤) لمعرفة ما إذا كان مجتمع ما يفضل نوعاً أ من سلعة ما ( معجون أسنان مثلا ) على نوع آخر ب أجرى استفتاء على عينة من ٣٠٠ شخص فوجد أن ١٦٨ شخصاً منهم يفضلون النوع أ وأن ١٣٢ شخصاً يفضلون النوع ب . فهل هذه النتيجة تعني أن المجتمع يفضل النوع أ ؟

(٥) في ٣٦٠ رمية لزهرتين من النرد ظهر ما مجموعة سبعة ٧٤ مرة وظهر ما مجموعة لإحدى عشر ٢٤ مرة . اختبر الفرض أن الزهرتين غير متحيزتين .

(٦) في المثال (٣ - ٩) عن توزيع خلايا الخميرة وفقنا توزيع بواسون لتوزيع تكرارى مشاهد في عينة . استخدم المستوى ٠,٠٥ لاختبار حسن المطابقة .

(٧) أخذ جوالان من حبوب الشعير لاختبار تأثير معالجة حرارية معينة على حيوية الحبوب وقد ترك الجوال أ دون معالجة ( مجموعة مراقبة ) وأعطيت المعالجة للجوال ب ثم أخذت عينة من ٨٠ حبة من كل جوال وفحصت بطريقة ما من حيث حيوية الحبوب فكانت النتيجة كما يلي :

لا يحتوى على	يحتوى على	
مقومات الحياة	مقومات الحياة	
١٦	٦٤	الجوال أ :
٤٦	٣٤	الجوال ب :

هل هناك فرق ذو دلالة بين الجوالين نتيجة للمعالجة الحرارية ؟ أى هل حيوية الحبوب مستقلة عن المعالجة الحرارية ؟

(٨) أخذت مجموعتان متكافئتان أ ، ب كل منهما تتكون من ١٠٠ شخص مريض بمرض معين وأعطى دواء للمجموعة الأولى ، ولم يعط للمجموعة الثانية ( مجموعة المراقبة ) فوجد أن عدد الذين شفوا من المجموعتين أ ، ب هما ٧٥ ، ٦٥ على الترتيب . أختبر الفرض أن هذا الدواء يساعد على الشفاء من المرض .

(٩) في عينة من خمس فراشات وجد أن تباين طول الجناح ١٣,٥٢ . إذا فرض أن مجتمع أجنحة هذه الفراشات معتدل ( بالنسبة لطول الجناح ) فأوجد فترة ثقة بدرجة ٩٥٪ لتباين هذا المجتمع .

(١٠) في دراسة عن نوع من الضفادع وجد في عينة من ٣٩ ضفدعا أن متوسط فترة نداءات الذكور ١٨٩ وحدة بانحراف معيارى ٣٢ وحدة .

(أ) أوجد فترة ثقة بدرجة ٩٥٪ لمتوسط فترة النداء في المجتمع .

(ب) أوجد فترة ثقة بدرجة ٩٥٪ لتباين فترة النداء في المجتمع .

(١١) بمجدول الاقتران الآتى التوزيع المشترك لمجموعة من ٩٥٠ طالبا من حيث خاصتى الذكاء والتغذية . هل تشير هذه البيانات إلى أن ذكاء الطلاب مستقل عن التغذية ؟

نوع التغذية	نسبة الذكاء				المجموع
	أقل من ٨٠	٨٠-٨٩	٩٠-٩٩	١٠٠ فأكثر	
تغذية حسنة	٢٤٥	٢٢٨	١٧٧	٢١٩	٨٦٩
سوء تغذية	٣١	٢٧	١٣	١٠	٨١
المجموع	٢٧٦	٢٥٥	١٩٠	٢٢٩	٩٥٠

## (٦ - ٧ - ٥) خاصة الجمع لتوزيعات $\chi^2$ المستقلة :

يمكن إثبات النظرية الآتية : « إذا كان لمتغير ما توزيع  $\chi^2$  بدرجات حرية  $\nu$  ، وكان لمتغير آخر مستقل عن المتغير الأول توزيع  $\chi^2$  بدرجات حرية  $\nu_1$  ، فإن المتغير الذى ينشأ من ضم هذين المتغيرين معا يكون له توزيع  $\chi^2$  بدرجات حرية  $\nu + \nu_1$  » وتعمد هذه النظرية لأى عدد من المتغيرات المستقلة .

### مثال (٦ - ١٣) :

فى دراسة عن تأثير التطعيم بمصل ما فى الوقاية من مرض ما أخذت مجموعتان متكافئتان من الأطفال طعمت إحداهما بالمصل ولم تطعم الأخرى . وبعد فترة محددة حسب عدد الذين أصيبوا والذين لم يصابوا بالمرض فى كلتا المجموعتين ووضعت النتائج فى جدول  $2 \times 2$  ووجد أن  $\chi^2 = 4,1$  بدرجة حرية واحدة . كررت نفس الدراسة مرتين أخريتين بشكل مستقل ( فى مكانين آخرين ) ووجد أن قيمتى  $\chi^2$  هما ٣,٩ ، ٤,٢ بدرجة حرية واحدة لكل منهما . نلاحظ ما يلى :

( أولا ) إذا أخذنا كلا من هذه التجارب الثلاث على حدة نجد أن كلا منها يشير إلى فرق ذى دلالة بين المجموعة التى طعمت بالمصل والمجموعة الضابطة عند المستوى ٠,٠٥ وذلك لأن القيمة الحرجة  $\chi^2_{0,05} = 3,841$  وهذا يعنى أن المصل كان له تأثير فى الحماية من المرض .

( ثانيا ) إذا ضمنا نتائج التجارب الثلاث معا ، وهذا جائز لأنها مستقلة ، نجد أن :

المجموع  $4,1 + 3,9 + 4,2 = 12,2$  بدرجات حرية عددها ثلاث ونظرا لأن القيمة الحرجة  $\chi^2_{0,05} = 7,815$  فإن هذه النتيجة تدعم كلا من النتائج الثلاث السابقة وتؤكد أن الفروق بين أزواج المجموعات لم تكن بالصدفة .

## THE F - TEST

## (٦ - ٨) اختبار ف :

إن الأهمية الرئيسية لهذا الاختبار هو قدرته على اختبار تساوى تبايني مجتمعين معتدلين  $\sigma_1^2$  ،  $\sigma_2^2$  عن طريق تبايني عيتين عشوائيتين مستقلتين  $\bar{X}_1$  ،  $\bar{X}_2$  مأخوذتين منهما . وقد رأينا أن اختبارات عن دلالة الفرق بين متوسطى عيتين بالصورة المبينة بالبند (٦ - ٣) يستلزم التأكد من تساوى تبايني المجتمعين ، ولذا ينبغي إجراء اختبار ف بشكل روتيني قبل تطبيق هذا الاختبار . ومن التطبيقات الهامة التي يستخدم فيها اختبار ف ذلك التطبيق المسمى بتحليل التباين الذى سنتناوله في الفصل الثامن حيث نحتاج أيضاً إلى اختبار تساوى تباينات المجتمعات .

ويبنى هذا الاختبار على الإحصاء :

$$(١٨) \quad \frac{\bar{X}_1^2}{\bar{X}_2^2} = \dots$$

التي تسمى بنسبة التباين variance ratio والتي يتطابق توزيعها مع توزيع المتغير ف السابق دراسته بلرجمتي حرية  $\nu_1 = 1$  ،  $\nu_2 = 1 - \nu_1$  تحت الافتراضات الآتية :

- (أ) أن تكون كل من العيتين عشوائية بسيطة .
- (ب) أن تكون العيتان مستقلتين .
- (ج) أن تكون العيتان مأخوذتين من مجتمع معتدل واحد أو من مجتمعين معتدلين لهما نفس التباين .

ملاحظة :

لاستخدام جدول ف للقيم الخارجة يجب أن توضع نسبة التباين بحيث تكون

أكبر من الواحد أى بحيث تكون  $\sigma^2 < \sigma^2$  مع ملاحظة أن  $\sigma^2$  تحدد العمود ،  
 $\sigma^2$  تحدد الصف .

مثال (٦ - ١٤) :

أخذت عيتان عشوائيتان حجمهما ٥ ، ٦ من الإناث والذكور من حشرة  
 ما وحسبت مدة البقاء على الحياة لكل حشرة بعد حرمانها من الطعام والشراب  
 فوجد أن تباين الأولى ٢٥,٧ وتباين الثانية ٠٧,٠٦ على أساس أن المجتمعين  
 معتدلين :

( أولا ) اختبر عند مستوى الدلالة ٠,٠١ أن تباين مجتمع الإناث أكبر من تباين  
 مجتمع الذكور .

( ثانيا ) اختبر عند مستوى الدلالة ٠,٠٥ أن المجتمعين متساويا التباين .

الحل :

$$f = \frac{25,7}{7,06} = 3,64 \text{ بدرجة حرية } 4, 5$$

( أولا ) الفرض الصفري  $F: \sigma^2 = \sigma^2$

الفرض البديل  $F: \sigma^2 < \sigma^2$  ( اختبار ذو جانب واحد )

من الجدول (٩) :  $F_{0,01} = 11,4$

بما أن  $3,64 < 11,4$  لا نستطيع رفض  $F$  عند المستوى ٠,٠١

( ثانياً ) الفرض الصفري  $F: \sigma^2 = \sigma^2$

الفرض الآخر  $F: \sigma^2 \neq \sigma^2$  ( اختبار ذو جانبيين مساحة كل منهما

(٠,٠٢٥

نريد إيجاد  $F_1$  ،  $F_2$  حيث  $L(F < F_1) = 0.025$  ،  
 $L(F < F_2) = 0.975$  مع ملاحظة أن الاحتمال  $0.975$  لا يوجد  
 بالجدول .

من الجدول  $F_{0.025, 1, 10} = 7.39$   
 إذن القيمة الحرجة اليمنى  $F_1 = 7.39$   
 المتساوية ل  $L(F < F_2) = 0.975$  تكافئ  
 المتساوية ل  $(\frac{1}{F_2} < \frac{1}{F_1})$   
 $0.025 =$



الشكل (٦-٤)

من الجدول  $F_{0.025, 1, 10} = 9.36$  راجع ملاحظة البند (٥-٣)

$$\therefore \frac{1}{F_2} = 9.36$$

$\therefore$  القيمة الحرجة اليسرى  $F_2 = 1 \div 9.36 = 0.107$   
 بما أن  $0.107 < 3.64 < 73.9$  لا نستطيع رفض  $F_0$  عند المستوى  $0.05$   
 أى أننا نقبل أن  $\sigma_1 = \sigma_2$  عند هذا المستوى .

## (٦ - ٩) فترات الثقة للنسبة في مجتمع :

نفرض أن نسبة وقوع حدث أ في مجتمع ما هو عدد مجهول ح ، فمثلا قد تعبر ح عن نسبة ظهور زهور حمراء من نوع من البلور ، أو نسبة الإصابة بمرض ما في نوع من الماشية ، أو نسبة النيكوتين في نوع من السجائر . يمكننا تقدير النسبة ح من واقع عينة عشوائية كبيرة وذلك بأخذ النسبة المشاهدة ر بين عدد مرات وقوع الحدث أ وعدد وحدات العينة ، وتزداد ثقتنا في التقدير كلما زاد حجم العينة . وإذا أردنا إيجاد صيغة لفترات الثقة للنسبة ح فإننا نستخدم التقدير ك كأساس لبناء هذه الصيغة وبحسب المنطق الآتي :

نعلم من البند (٣ - ٣) أنه تحت شروط عشوائية العينة وثبات النسبة ح واستقلال الأحداث يكون للمتغير العشوائي  $\sqrt{\frac{r}{n}}$  الذي يعبر عن عدد مرات وقوع الحدث أ في العينات ذوات الحجم ن توزيع ذي الحدين دليلا ن ، ح وسطه الحسابي ن ح وتباينه ن ح ك حيث ك = ١ - ح . ونعلم من البند (٦ - ٤) أنه إذا كان المتغير العشوائي  $\sqrt{\frac{r}{n}}$  يعبر عن النسبة المشاهدة في العينة فإن توزيع الإحصاءة :

$$\frac{r - n}{\sqrt{\frac{r}{n}}} = \sqrt{n}$$

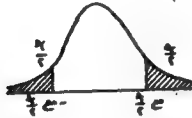
يقترّب من التوزيع المعتدل المعياري بشرط أن تكون ن كبيرة وألا تكون ح أو ك قريبة من الصفر .

في هذه الحالة يمكن إثبات أن الفترة :

$$(19) \quad \left( \sqrt{\frac{(n-1)r}{n}} \right) \frac{1}{\sqrt{n}} + \sqrt{\frac{(n-1)r}{n}} \frac{1}{\sqrt{n}} , \left( \sqrt{\frac{(n-1)r}{n}} \right) \frac{1}{\sqrt{n}} - \sqrt{\frac{(n-1)r}{n}} \frac{1}{\sqrt{n}}$$

هى فترة ثقة بدرجة  $(1 - \alpha)$  للنسبة ح .  
 حيث مع  $\frac{\alpha}{2}$  هى القيمة الحرجة في التوزيع المعتدل المعيارى التي تحقق المعادلة

$$L \left( \text{مع} \frac{\alpha}{2} < \varepsilon < \text{مع} \frac{\alpha}{2} \right) = 1 - \alpha \quad (20)$$



الشكل (٦-٥)

فمثلا حين  $\alpha = 0.05$  فإن مع  $\frac{\alpha}{2} = 0.025$  راجع المثال (٤-٣) .  
 وحين  $\alpha = 0.01$  فإن مع  $\frac{\alpha}{2} = 0.005$  راجع المثال (٤-٣) .  
 ويلاحظ أن أى عدد يقع في هذه الفترة يصلح لأن يؤخذ كتقدير للنسبة ح .

مثال (٦-١٥) :

أخذت عينة عشوائية حجمها ٤٠٠ شخص من مجتمع ما ووجد أن ٢٣ منهم مصابون بمرض البول السكرى . أوجد فترة ثقة بدرجة ٩٥٪ لنسبة المصابين بهذا المرض في هذا المجتمع .

الحل :

$$s = \frac{23}{400} = 0.0575 \quad \text{نسبة الإصابة في العينة}$$

نأخذ هذه النسبة كتقدير للنسبة ح مع ملاحظة عشوائية العينة وكبر حجمها .

$$\text{تقدير الخطأ المعياري} = \frac{(s-1)s}{n} = \frac{(0.0575-1)0.0575}{400} = 0.0116$$



من (١٩) وبأخذ  $\alpha = ٠,٠٥$  نجد أن  
 الحد الأدنى للفترة  $= ٠,٠٥٧٥ - ٠,٠١١٦ \times ١,٩٦ = ٠,٠٣٥$   
 الحد الأعلى للفترة  $= ٠,٠٥٧٥ + ٠,٠١١٦ \times ١,٩٦ = ٠,٠٨٠٢$   
 إذن الفترة (٠,٠٣٥ ، ٠,٠٨٠) هي فترة ثقة بدرجة ٩٥٪ لنسبة المصابين بالبول  
 السكري في المجتمع .

### (٦ - ٩ - ١) اختبار دلالة الفرق بين نسبتين عيتين مستقلتين :

نفرض أن لدينا مجتمعين نسبة وقوع حدث ما في أحدهما  $ح_١$  ونسبة وقوعه  
 في الآخر  $ح_٢$  . ونفرض أننا حصلنا على عيتين عشوائيتين مستقلتين من هذين  
 المجتمعين ووجدنا أن نسبتين وقوع الحدث فيهما  $ر_١$  ،  $ر_٢$  على الترتيب . يمكن  
 إثبات أن توزيع المعاينة للإحصاءة .

$$(٢١) \quad \bar{u} = \frac{(\bar{r}_2 - \bar{r}_1) - (\bar{p}_2 - \bar{p}_1)}{\sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2}}}$$

يقرب من التوزيع المعتدل المعياري كلما كبرت كل من  $n_١$  ،  $n_٢$  ولم تكن أي  
 من  $ح_١$  ،  $ك_١$  ،  $ح_٢$  ،  $ك_٢$  قريبة من الصفر .  
 وإذا فرضنا أن  $ح_١ = ح_٢ = ح$  ،  $ك_١ = ١ - ح$  فإن هذه الإحصاءة تكتب .

$$(٢١) \quad \bar{u} = \frac{\bar{r}_2 - \bar{r}_1}{\sqrt{ح \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

وفي هذه الحالة تقدر  $ح$  من العيتين كالآتي :

$$\bar{ح} = \frac{n_2 \bar{r}_1 + n_1 \bar{r}_2}{n_1 + n_2}$$

وعلى ذلك يمكن استخدام الإحصاءة (٢١) للكشف عما إذا كان الفرق المشاهد بين  $s_1$  ،  $s_2$  في العيتين هو فرق صغير لا دلالة له أم فرق يدل على وجود فرق حقيقي بين النسبتين  $H_0$  ،  $H_1$  في المجتمعين .

مثال (٦ - ١٦) :

لمعرفة تأثير طعم ما في الوقاية من وباء الكوليرا اختبرت عيتان عشوائيتان حجم الأولى ١٠٠٠ شخص وحجم الثانية ١٥٠٠ شخص وحقت أفراد المجموعة الأولى بالطعم ولم يحقن أفراد المجموعة الثانية ( مجموعة مراقبة ) . وبعد فترة من الزمن ظهرت ١٠٠ حالة مرضية في المجموعة الأولى و ٥٠٠ حالة في المجموعة الثانية . اختبر ما إذا كان لهذا الطعم أثر في الوقاية من هذا المرض مستخدماً مستوى الدلالة ٠,٠١

الحل :

الفرض الصفري  $F_0$  :  $H_0 = H_1$  ( لا يوجد تأثير للطعم المعطى )  
الفرض الآخر  $F_1$  :  $H_0 > H_1$  ( اختبار ذو جانب واحد ) .

$$\text{نسبة الإصابة في المجموعة الأولى } s_1 = \frac{100}{1000} = 0,10$$

$$\text{نسبة الإصابة في المجموعة الثانية } s_2 = \frac{500}{1500} = 0,33$$

$$\text{من (٢١) : تقدير } H = \frac{1500 \times 0,33 + 1000 \times 0,1}{1500 + 1000} = 0,238$$

$$\text{تقدير الخطأ المعياري} = \sqrt{H \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{0,238 \times \left( \frac{1}{1000} + \frac{1}{1500} \right)} = 0,017$$

$$\text{من (٢١) : صـحـه} = \frac{0,33 - 0,10}{0,017} = 13,52$$

بما أن  $13,52 > 2,33$  نرفض الفرض الصفرى عند المستوى  $0,01$  ونستنتج أن للطعم تأثير في الوقاية من المرض .

### (٦ - ٩ - ٢) فترات الثقة للفرق بين نسبتي مجتمعين :

يمكن إثبات أنه تحت الشروط سابقة الذكر تكون الفترة :

$$(22) \quad \left\{ \frac{(x_1 - y_1) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \leq Z \leq \frac{(x_1 - y_1) + (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \right\}$$

حيث  $x, y$  :  $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

هى فترة ثقة بدرجة  $(1 - \alpha)$  للفرق  $p_1 - p_2$  بين نسبتي المجتمعين . ففي المثال (٦ - ١٥) الأخير نجد أن :

$$\text{الحد الأدنى للفترة} = 2,58 \times 0,017 - (0,1 - 0,33) = 0,186$$

$$\text{الحد الأعلى للفترة} = 2,58 \times 0,017 + (0,1 - 0,33) = 0,274$$

أى أن الفترة  $(0,186, 0,274)$  هى فترة ثقة بدرجة ٩٩٪ للفرق بين النسبتين في المجتمعين .

### تمارين (٦ - ٣)

(١) في عينة عشوائية من ٤٠٠ شخص حققوا بمصل ما تأثر ١٣٦ تأثراً ضاراً . أوجد فترة ثقة بدرجة ٩٥٪ لنسبة الأشخاص الذين يتأثرون تأثراً ضاراً من هذا المصل .

(٢) من ١٠٠ سمكة صيدت من بحيرة ما وجد أن ١٦ منها لا تصلح للأكل نتيجة لتلوث كيميائي في يفة هذه البحيرة . أنشئ فترة ثقة بدرجة ٩٥٪ لاحتمال أنه إذا صيدت سمكة من هذه البحيرة تكون غير صالحة للأكل .

(٣) في تجربة عن تهجين فيران ذوات الشعر النحيف المتهدل وفيران ذوات الشعر المجعد وجد في ٣٢ ولادة بكل منها ٨ فيران أن عدد الولادات التي تحتوي بالضبط على ٣ فأراً ذا شعر نحيف ومتهدل كآلاتي :

س : ٠ ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨  
عدد الولادات : ٠ ١ ٢ ٤ ١٢ ٥ ٢ ٠

على فرض أن توزيع ذى الحدين يصلح نموذجاً في هذه التجربة أوجد فترة ثقة بدرجة ٩٩٪ لاحتلال الحدث « ولادة فأر ذى شعر نحيف ومتهدل » - راجع البند (٣ - ٣ - ١) .

(٤) أخذت مجموعتان عشوائيتان بكل منهما ٨٠ مريضاً . أعطى للمجموعة الأولى أقراص تحتوي على دواء ضد الحساسية وأعطى للمجموعة الثانية أقراص تمويه ( لا تحتوي على أى دواء ) ، فظهرت أعراض الحساسية في ٢٣ من المجموعة الأولى ، ٥١ من المجموعة الثانية . اختبر عند مستوى الدلالة ٠,٠١ ما إذا كانت نسبة ظهور الحساسية في المرضى الذين يتلقون أقراص الحساسية لا تختلف عنها في المرضى الذين يتلقون أقراص التمويه .

(٥) في عينة عشوائية من ٢٠٠ شخص لا يتناولون طعام الإفطار أفاد ٨٢ منهم أنهم يشعرون بتعب منتصف الصباح . وفي عينة عشوائية من ٤٠٠ شخص يتناولون طعام الإفطار أفاد ١١٦ منهم أنهم يشعرون بتعب منتصف الصباح . استخدم مستوى الدلالة ٠,٠١ لاختبار الفرض الصفري أنه لا يوجد فرق بين المجتمعين ضد الفرض الآخر أن تعب منتصف الصباح أكثر تفشيًا بين الأشخاص الذين لا يتناولون طعام الإفطار .

## (٦ - ١٠) تحديد حجم العينة :

من الأمور التي تشغل الباحث عند تصميم تجربة لحل مشكلة ما تحديد عدد

وحدات العينة ( العشوائية ) اللازم لضمان أن تكون أحكامه عن المجتمع الذى يدرسه على درجة كافية من العمومية والدقة . وبطبيعة الحال كلما كانت العينة كبيرة كلما زادت الثقة في هذه الأحكام ، غير أن كبر حجم العينة يحتاج إلى الكثير من الجهد والوقت والتكاليف ، سواء في عملية المعاينة أو في قياس وتحليل البيانات ولذلك فإن كفاءة التصميم تتطلب إيجاد حد أعلى معقول لحجم العينة . وسنبحث هذا الموضوع في الحالات الثلاث الآتية .

( أولا ) عند تقدير الوسط الحسابى لمجتمع :

نفرض أننا نتساءل عن حجم العينة المناسب لتقدير متوسط المجتمع  $\mu$  من المتوسط  $\bar{x}$  لعينة عشوائية مأخوذة منه . إن الإجابة عن هذا التساؤل لا تتأتى إلا إذا أجبنا عن السؤال العملى الآتى : « ما مقدار الخطأ الذى يمكن السماح به عند تقدير  $\mu$  عن طريق  $\bar{x}$  ؟ أى ما هو الحد الأعلى الذى يمكن التجاوز عنه لانحراف  $\bar{x}$  عن القيمة الحقيقية  $\mu$  ؟ » وهذا السؤال لا يجاب عنه إحصائيا وإنما هو من اختصاص الباحث التطبيقى وهو الذى يجيب عنه من واقع خبرته بميدان البحث . فإذا رأى الباحث أن الحد الأعلى للخطأ المسموح به هو عدد ما خ ، ورأى في الوقت نفسه أن يعين درجة ثقة (٩٥٪ مثلا ) في عدم تخطى هذا الحد عند التطبيق ، فإن الحجم المناسب للعينة الذى يحقق الفرض المنشود ينتج حسب الأساس الآتى :

من البند (٦ - ٣) نعلم أنه إذا كان لدينا مجتمع معتدل وسطه الحسابى  $\mu$  وانحرافه المعيارى  $\sigma$  فإن الإحصاءة  $\bar{x}$  للعينات ذوات الحجم  $n$  يكون لها توزيع معتدل وسطه الحسابى  $\mu$  وانحرافه المعيارى  $\frac{\sigma}{\sqrt{n}}$  وتحقق هذه النظرية أيضا ( بالتقريب ) حين لا يكون المجتمع معتدلا بشرط أن يكون حجم العينة كبيرا  
(  $n \geq 30$  ) .

وإذا كانت  $\bar{x}$  هي الوسط الحسابي لعينة عشوائية ما فإن الفترة :

$$(\bar{x} - \frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\sigma}{\sqrt{n}} , \bar{x} + \frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\sigma}{\sqrt{n}})$$

تكون فترة ثقة بدرجة  $1 - \alpha$  للمتوسط  $\mu$  للمجتمع ، حيث مع  $\frac{\sigma}{\sqrt{n}}$  هي قيمة المتغير المعتدل المعيارى التي تحقق المعادلة :

$$L(\text{مع } \frac{\sigma}{\sqrt{n}} < E < \text{مع } \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

وبذلك تكون أكبر قيمة للانحراف  $|\mu - \bar{x}|$  هي  $\frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\sigma}{\sqrt{n}}$  ويسمى هذا العدد بمد الخطأ error bound ، بدرجة ثقة  $1 - \alpha$

وإذا اخترنا أن يكون الخطأ المسموح به هو مقدار معين  $x$  وأردنا أن نكون على ثقة بدرجة  $(1 - \alpha)$  بألا يتعدى الخطأ الذى تقع فيه القيمة  $x$  فإن حجم

العينة المطلوب يجب أن يحقق المعادلة  $\frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\sigma}{\sqrt{n}} = x$  . وبحل هذه المعادلة في  $n$  نجد أن :

$$(23) \quad \left( \frac{\frac{\sigma}{\sqrt{n}} \text{ مع } \frac{\sigma}{\sqrt{n}}}{x} \right)^2 = n$$

وهذه هي القيمة المطلوبة لحجم العينة  $n$  الذى يكفي لتحقيق الغرض المطلوب . كما يمكن أن نأخذ العدد  $n$  حيث

$$(24) \quad \frac{\sigma}{\sqrt{x} \alpha} = n$$

كحد أعلى لحجم العينة . وتنتج هذه المعادلة من متباينة تشيشف الشهيرة التي لا تتطلب توزيعاً أو شروطاً معينة ، إلا أنها غالباً ما تعطى قيمة أكبر مما يبنى لحجم العينة .

ويلاحظ أن إيجاد قيمة  $n$  من أى من المعادلتين (٢٣) أو (٢٤) يتطلب معرفة ولو تقريبية بالانحراف المعياري  $\sigma$  للمجتمع . وحين تكون قيمة  $\sigma$  مجهولة تماماً فلا مفر من تقديرها من عينة عشوائية استطلاعية كبيرة لا يقل حجمها عن ٣٠ .

مثال (٦ - ١٧) :

في عينة عشوائية حجمها ٣٦ وجد أن الوسط الحسابي والانحراف المعياري هما ٢,٦ و ٠,٣ على الترتيب . ما حجم العينة اللازم لإعطائنا ثقة بدرجة ٩٥٪ بألا يزيد الخطأ في تقدير الوسط الحسابي  $\mu$  للمجتمع عن ٠,٠٦ ؟

الحل :

لدينا  $x = ٠,٠٦$  ،  $١ - \alpha = ٠,٩٥$  ومنها  $\alpha = ٠,٠٥$  ،

$$مع = \frac{1,96}{\sqrt{36}}$$

نأخذ الانحراف المعياري ٠,٣ الناتج من العينة الاستطلاعية كتقدير للانحراف المعياري للمجتمع . ثم نحسب  $n$  من الصيغة (٢٣) كالآتي :

$$n = \left( \frac{1,96 \times ٠,٣}{٠,٠٦} \right)^2 = ٩٦,٠٤$$

إذن الحجم المطلوب للعينة هو ٩٧ أى أنه يكفينا أن نوجد الوسط الحسابي  $\bar{x}$  لعينة عشوائية من هذا الحجم لكي نكون واثقين بدرجة ٩٥٪ أن اختلاف  $\bar{x}$  عن  $\mu$  لا يزيد عن ٠,٠٦ .

نلاحظ أنه يأخذ  $n = 97$  فإن  $\frac{E}{\sqrt{n}} = 1,96 \times \frac{E}{\sqrt{n}} = \frac{1,96 \times 0,3}{\sqrt{97}} = 0,097$  وهذا أصغر من  $0,06$ .

(ثانياً) عند تقدير الفرق بين متوسطي مجتمعين معتلين :

بنفس الطريقة يمكن إيجاد حد أعلى لحجم العينة اللازم لتقدير الفرق  $\mu - \mu$  بين متوسطي مجتمعين معتلين بحيث نكون على ثقة بدرجة  $1 - \alpha$  بالآ يزيد المقدار  $\bar{x} - \bar{y}$  المشاهد في عيتين عن المقدار  $\mu - \mu$  للمجتمعين ، عن قيمة معينة  $x$  مفروضة مسبقاً ، وما علينا إلا استخدام الخطأ المعياري لتوزيع المعاينة للإحصاءة  $\bar{x} - \bar{y}$  وهو  $\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}}$  بدلا من الخطأ المعياري  $\frac{\sigma}{\sqrt{n}}$  لتوزيع المعاينة للإحصاءة  $\bar{x} - \bar{y}$  . وإذا كانت العيتان متساويتى

الحجم :  $n = n = n$  وكان المجتمعان متساويين في التباين :  $\sigma^2 = \sigma^2 = \sigma^2$  فإن ذلك الخطأ المعياري يأخذ الصورة  $\frac{\sqrt{\sigma^2 2}}{\sqrt{n}}$  ويكون الحد الأعلى المطلوب لحجم

$$\text{كل من العيتين هو} \quad \left( \frac{\frac{\alpha}{2} \sigma \sqrt{2}}{x} \right)^2 = n \quad (25)$$

ويلاحظ أن هذه الصيغة ضعف الصيغة (٢٣) .

مثال (٦ - ١٨) :

في تجربة تهدف إلى اختبار دواء تخسيس وتحكم في زيادة الوزن عند النساء ، رؤى البدء باستخدام الدواء على إناث القطط المنزلية ، فاختيرت عينة عشوائية من



٦٠ قطة قسمت عشوائيا إلى مجموعتين متكافئتين في الوزن بكل منهما ٣٠ قطة ، مجموعة تجريبية وأخرى ضابطة ، أعطى الدواء إلى المجموعة التجريبية فقط ووضعت المجموعتان تحت نفس الظروف . وبعد ٦ أسابيع وباستخدام وحدة قياس معينة وجد أن متوسط النقص في وزن المجموعة التجريبية ١٠ وحدات ومتوسط النقص في وزن المجموعة الضابطة ١٣,٣ وحدة . على فرض أن تباين كل من المجموعتين ٤٠ أوجد حدا أعلى لحجم كل من العيتين الذي يعطينا ثقة بدرجة ٩٥٪ ألا يزيد الفرق المشاهد في متوسطي العيتين عن الفرق الحقيقي بين متوسطي المجموعتين عن ٣ وحدات .

الحل :

لدينا  $\sigma^2 = 40$  ،  $\alpha = 0.05$  ،  $\pm z_{\alpha/2} = 1.96$  ، وإذا  $\frac{\sigma^2}{n} = 1.96$

من الصيغة (٢٥) :  $n = \frac{40 \times 2}{1.96^2} = 24.148 \sim 25$

( ثالثا ) عند تقدير نسبة وقوع حدث في مجتمع :

نفرض أن نسبة وقوع حدث معين في مجتمع هو مقدار ثابت مجهول  $\pi$  ونفرض أننا نرغب في معرفة حجم العينة المناسب لتقدير هذه النسبة عن طريق النسبة  $\hat{\pi}$  التي تظهر في عينة عشوائية ، بحيث نكون على ثقة بدرجة  $1 - \alpha$  ألا يزيد الخطأ الناشئ عن هذا التقدير عن مقدار معين  $\pi$  .

من البند (٦ - ٤) نعلم أنه إذا كان حجم العينة كبيرا  $(n \geq 30)$  فإن توزيع المعاينة للنسبة  $\hat{\pi}$  لوقوع هذا الحدث في العينات ذوات الحجم  $n$  يقترب من توزيع معتدل وسطه الحسابي  $\pi$  وانحرافه المعياري  $\sqrt{\frac{\pi(1-\pi)}{n}}$  حيث  $k = 1 - \alpha$  . في

هذه الحالة يكون حد الخطأ أى أكبر قيمة للمقدار  $|\hat{\pi} - \pi|$  هو  $\sqrt{\frac{\pi(1-\pi)}{n}}$  مع



## مثال (٦ - ١٩) :

في عملية مسح صحي عن طريق العينة ، يراد تقدير النسبة ح للأشخاص ضعاف البصر . كم شخصاً ينبغي اختبارهم إذا كان المسئولون يريدون أن يتأكدوا بدرجة ٩٨٪ أن الخطأ في التقدير يقع في المدى  $\pm ٠,٠٥$  .

(أ) في حالة عدم وجود أى معرفة لقيمة ح

(ب) إذا كان المعروف أن ح = ٠,٣ بالتقريب ؟

## الحل :

لدينا  $\alpha = ٠,٠٥$  ،  $\alpha - ١ = ٠,٩٨$  إذن  $\alpha = ٠,٠٢$  ،  $\alpha = ٠,٠١$  القيمة هي قيمة المتغير المعتدل المعيارى أ التى تحقق المعادلة

$$ل (أ) < \epsilon < ١ - \alpha = ٠,٩٨$$

$$ل (أ) < \epsilon < ١ - \alpha = ٠,٩٨ \div ٢ = ٠,٤٩$$

من جدول المساحات أسفل المنحني المعتدل نجد أن  $أ = ٢,٣٣$

$$(أ) من (٢٧) : \frac{١}{٤} = \frac{٢,٣٣}{٠,٠٥} \Rightarrow ٠,٤٢,٨٩$$

إذن الحجم المطلوب للعينة هو ٥٤٣ .

$$(ب) من (٢٨) : ٠,٣ \times ٠,٧ = \frac{٢,٣٣}{٠,٠٥} \Rightarrow ٤٥٦,٠٢٧٦$$

إذن الحجم المطلوب للعينة هو ٤٥٦ .

## تمارين (٦ - ٤)

(١) يريد باحث تقدير متوسط محتوى الفسفات في وحدة حجم من ماء إحدى البحيرات والمعروف من دراسات سابقة أن الانحراف المعيارى لهذا المحتوى ذو قيمة

تكاد تكون ثابتة عند  $\sigma = 4$  . ما عدد عينات الماء التي ينبغي للباحث تحليلها لكي يثق بدرجة ٩٠٪ أن الخطأ في التقدير لا يتعدى ٠,٨ ؟

(٢) أ - أراد مهندس تقدير الوسط الحسابي للمدة التي تجف فيها خلطة معينة من الأسمنت تستخدم في إصلاح الطرق . وقد جرب هذه الخلطة في ١٠٠ بقعة ووجد أن الوسط الحسابي والانحراف المعياري لمدة الجفاف هما ٣٢ دقيقة و ٤ دقائق على الترتيب . استخدم هذه المعلومات لإيجاد فترة ثقة بدرجة ٩٩٪ لمتوسط مدة جفاف هذه الخلطة .

ب - أراد المهندس أن يحدد حجم العينة ( أى عدد البقع التي يجرب فيها الخلطة ) لكي يكون متأكداً بدرجة ٩٥٪ أن متوسط مدة الجفاف المحسوب من عينة بهذا الحجم لا يختلف عن المتوسط الحقيقي لهذه المدة إلا بدقيقة واحدة على الأكثر . أوجد هذا الحجم علماً بأن الخبرة السابقة تشير إلى أن الانحراف المعياري لمدة الجفاف هو ٥ دقائق بالتقريب .

(٣) في أحد مراكز القلب والصدر يراد تقدير المعدل ح لوقوع حالات ضيق التنفس بين الذكور من متوسطى العمر ممن كانوا يدخنون أكثر من علبتين من السجائر في اليوم خلال خمس سنوات سابقة . ما حجم العينة التي نختاره بحيث نكون على ثقة بدرجة ٩٥٪ أن الخطأ في تقدير النسبة ح لا يزيد عن ٠,٠٣ ، علماً بأن المعروف أن قيمة ح الحقيقية قريبة من ٠,١٥ ؟

## QUALITY CONTROL (٩ - ١١) مراقبة الإنتاج

من تطبيقات نظرية العينات في الصناعة ذلك التطبيق المسمى بمراقبة جودة الإنتاج ومعاينات القبول ، وهو يتعلق بإحدى المشكلات التي تظهر في المصانع التي تنتج سلعا على نطاق واسع . ويتمثل هذا التطبيق في استخدام فكرة فترات الثقة في التفيش على جودة السلع المنتجة أثناء عملية الإنتاج .

ومن المعروف أن الوحدات المنتجة في أى مصنع لا يمكن أن تكون جميعها متشابهة تماماً مهما تقدمت التكنولوجيا الصناعية بل توجد دائماً انحرافات صغيرة عن المواصفات الموضوعية للسلعة تنشأ عن عدد كبير من العوامل الصغيرة التي لا يمكن التحكم فيها وتعتبر عوامل عشوائية ، ولا مفر للمصنع من أن يسمح بالتجاوز عن هذه الانحرافات طالما كانت في حدود معقولة يقبلها المنتج والمستهلك ، أما إذا خرجت الانحرافات عن هذه الحدود فإن المصنع يشتبه في وجود خلل ما إما في أجزاء آلة المصنع أو في سلوك العمال أو في الإدارة أو أية مصادر أخرى للأخطاء ، وعليه حيثئذ أن يتحرى عن أسباب هذا الخلل ويعمل على تلافيه . وليس من المعقول أن يفحص المصنع كل وحدة ينتجها والمتبع أن يجرى الآتى .

نفرض مثلاً أن مصنعاً ينتج يومياً عشرات الألوف من الأنابيب الاسطوانية ذات مواصفات معينة منها أن طول الاسطوانة ٦ سنتيمترات مثلاً . ليطمئن المصنع على توفر هذه الصفة يضع اختباراً إحصائياً ليختبر به الفرض أن الوحدات المنتجة تتوفر فيها الخاصية المطلوبة ، أى ليختبر الفرض الصفري  $\mu = 6$  ضد الفرض  $\mu \neq 6$  . ويجرى هذا الاختبار في فترات منتظمة من الزمن ، مثلاً كل نصف ساعة أو كل ساعة ، وفي كل مرة يأخذ عينة عادة من ٣ إلى ١٠ وحدات ويقيس متوسط الطول  $\bar{x}$  فيها و يطبق الاختبار . فإذا أدى الاختبار إلى قبول الفرض الصفري فإن هذا يعنى أن الإنتاج يسير سيراً طبيعياً ، أما إذا أدى إلى رفض الفرض الصفري فإن هذا يعنى أن انحرافات أطوال الوحدات المنتجة عن الطول المطلوب هي انحرافات جوهرية وينبئ حيثئذ التحرى عن أسباب هذه الانحرافات . وهذا ما يسمى بمراقبة الإنتاج .

وتوضع قاعدة الاختبار على هيئة فترة ثقة مركزها القيمة  $\mu$  المطلوبة ( أى أن حدى الثقة يكونان على بعدين متساويين من  $\mu$  ) . وإذا افترضنا أن توزيع الأطوال في الوحدات المنتجة معدل متوسطه  $\mu$  وانحرافه المعياري  $\sigma$  فإن توزيع المعاينة

للمتوسطات الحسابية للعينات التي من الحجم  $n$  يكون معتدلاً متوسطه  $\mu$  وانحرافه المعياري  $\frac{\sigma}{\sqrt{n}}$  - راجع البند (٦ - ٣ - أولاً) - وتكون الفترة

$$(\bar{x} - 2,08 \times \frac{\sigma}{\sqrt{n}}, \bar{x} + 2,08 \times \frac{\sigma}{\sqrt{n}})$$

هي فترة ثقة بدرجة ٩٩٪ للمتوسط  $\mu$ . أي أنه في ٩٩٪ من العينات التي نأخذها تقع  $\mu$  في هذه الفترة، أي تتحقق المتباينة الآتية:

$$\bar{x} - 2,08 \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2,08 \times \frac{\sigma}{\sqrt{n}}$$

ومن هذه المتباينة المزدوجة ينتج أن

$$\mu - 2,08 \times \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 2,08 \times \frac{\sigma}{\sqrt{n}} \quad (٢٩)$$

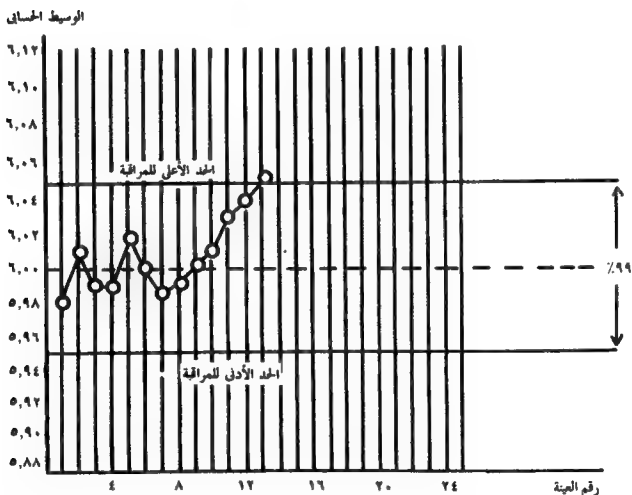
ويسمى الطرف الأيسر من هذه المتباينة بمحد المراقبة الأدنى، ويسمى الطرف الأيمن منها بمحد المراقبة الأعلى *upper and lower control limits* وعند المراقبة إذا أخذت عينة ووجد أن متوسطها  $\bar{x}$  يقع بين هذين الحدين يقبل الفرض الصفري ويستمر الإنتاج دون تعديل، أما إذا ساوت  $\bar{x}$  أحد الحدين أو تعدته فيرفض الفرض الصفري عند مستوى الدلالة ٠,٠١. وينبغي حينئذ اتخاذ ما يلزم (وربما إيقاف الآلة) لتحري أسباب الحلل وإصلاحه.

وفي المعتاد تستخدم فترة ثقة بدرجة ٩٩٪ أما إذا أردنا استخدام فترة ثقة بدرجة ٩٥٪ فإننا نضع العدد ١,٩٦ بدلاً من العدد ٢,٠٨ في المتباينة (٢٩) - راجع المثال (٤ - ٣) - أما قيمة  $\sigma$  فيمكن تقديرها من عينة عشوائية كبيرة تؤخذ من الإنتاج عندما يكون في حالته الطبيعية.

ولتسهيل عملية المراقبة ترصد قيمة المتوسط  $\bar{x}$  لكل عينة مختارة بالترتيب في خريطة تسمى بخريطة المراقبة control chart . فإذا فرضنا أن  $\mu = 6$  ،

$$\sigma = 0.6 ، \quad n = 10 \quad \text{فان حدى المراقبة يكونان} \quad 6 \pm \frac{0.6}{\sqrt{10}} \times 2.58$$

أى بالتقريب ٥,٩٥ ، ٦,٠٥ من الستيمترات وتكون خريطة المراقبة كالآتي :



الشكل (٦-٦) : خريطة المراقبة للوسط الحسابى

وتبين هذه الخريطة الخططين الممثلين للحددين الأدنى والأعلى للمراقبة ، كما تمثل النقط متوسطات ١٣ عينة (حجم كل منها ١٠) بترتيب اختيارها وهى  
 ٥,٩٨١ ، ٦,٠١١ ، ٥,٩٩١ ، ٥,٩٩٢ ، ٦,٠٢٢ ، ٦,٠٠٣ ، ٥,٩٨٥ ،  
 ٥,٩٩٢ ، ٦,٠٠٤ ، ٦,٠١٣ ، ٦,٠٣٢ ، ٦,٠٤٣ ، ٦,٠٥٤ ، ويبدو من  
 قيمة المتوسط الأخير أن الإنتاج فى حاجة إلى تعديل عند الوصول إلى العينة ١٣ .  
 هذا ويمكن بنفس الطريقة متابعة التطور فى تباين الإنتاج أو فى أى بارامتر آخر .

## تمارين (٦ - ٥)

أخذت عشر عينات حجم كل منها ٤ وحدات فى فترات منتظمة ووجد أن  
 سمك الوحدات كما فى الجدول الآتى . مثل متوسطات هذه العينات على خريطة  
 مراقبة على فرض أن المجتمع معتدل متوسطه ٥ وانحرافه المعيارى ١,٥٥ .

الزمن	٨/-	٨/٣٠	٩/-	٩/٣٠	١٠/-	١٠/٣٠	١١/-	١١/٣٠	١٢/-	١٢/٣٠
سمك الوحدة	٣	٣	٥	٧	٧	٤	٥	٦	٥	٥
	٤	٦	٢	٥	٣	٤	٦	٤	٥	٢
	٨	٦	٥	٤	٦	٣	٤	٦	٦	٥
	٤	٨	٦	٤	٥	٦	٦	٤	٤	٣



## الفصل السابع

### حساسية اختبارات الفروض

#### SENSITIVITY OF TESTS OF HYPOTHESES

لا يوجد قرار إحصائي منزه عن الخطأ ، فالقرارات الإحصائية هي دائماً قرارات احتمالية بمعنى أنه لا مفر من وجود احتمال للخطأ في أى قرار نصدره عن مجتمع عن طريق عينة . ولما كانت هذه القرارات مؤسسة على ما نجره من اختبارات للفروض ونزيد ثقتنا فيها بزيادة حساسية هذه الاختبارات ، وجب علينا أن ندرس كيف نزيد من هذه الحساسية أى من قدرة الاختبارات على تمكيننا من اتخاذ القرار السليم الذى لا يشوبه إلا قدر ضئيل من الخطأ . ويتأتى ذلك عن طريق التحكم ما أمكن في احتمالات الأخطاء التى تنجم حتما عند استخدام هذه الاختبارات . ومن الطبعى إذن أن نبدأ بتقديم هذه الأخطاء توطئة لدراسة كيفية التقليل منها ما استطعنا إلى ذلك سبيلا .

#### (٧ - ١) نوعا الأخطاء الإحصائية :

نعلم حتى الآن أننا حين نكون بصدد اتخاذ قرار برفض أو قبول فرض صفري ف ضد فرض آخر ف<sub>١</sub> عند مستوى معين  $\alpha$  من الدلالة ، نقوم أولاً بتجزئء فضاء العينة إلى منطقتين منفصلتين ومتكاملتين ١ ، ٢ نسمى إحداهما ١ بمنطقة الرفض أو بالمنطقة المخرجة ونسمى الأخرى ٢ بمنطقة القبول . وإذا وقعت قيمة مشاهدة من إحصاء الاختبار في المنطقة ٢ رفضنا الفرض الصفري عند المستوى

$\alpha$  لصالح الفرض الآخر ، أما إذا وقعت القيمة المشاهدة في المنطقة  $\bar{C}$  فإننا نقبل الفرض الصفري .

ونظراً لأننا لا نعرف مسبقاً ما إذا كان الفرض الصفري صحيحاً أو زائفاً فإن القرار الذي نتخذه يكون على إحدى الحالات الآتية :

- (١) أن يكون الفرض الصفري صحيحاً ويكون قرارنا هو رفض هذا الفرض .
- (٢) أن يكون الفرض الصفري صحيحاً ويكون قرارنا هو قبول هذا الفرض .
- (٣) أن يكون الفرض الصفري زائفاً ويكون قرارنا هو رفض هذا الفرض .
- (٤) أن يكون الفرض الصفري زائفاً ويكون قرارنا هو قبول هذا الفرض .

ومن الواضح أن قرارنا يكون صائباً في الحالتين (٢) و (٣) ويكون خاطئاً في الحالتين (١) و (٤) . يقال للخطأ الناشئ عن القرار (١) إنه خطأ من النوع الأول ، كما يقال للخطأ الناشئ عن القرار (٤) إنه خطأ من النوع الثاني . ويعرف هذان النوعان من الخطأ كالآتي :

### الخطأ من النوع الأول : TYPE I ERROR

هو ذلك الخطأ الذي ينشأ حين نتخذ قراراً برفض الفرض الصفري بينما يكون هذا الفرض صحيحاً في الواقع . ويرمز لاحتمال هذا الخطأ بالرمز  $\alpha$  .

### الخطأ من النوع الثاني : TYPE II ERROR

هو ذلك الخطأ الذي ينشأ حين نتخذ قراراً بقبول الفرض الصفري بينما يكون هذا الفرض زائفاً في الواقع . ويرمز لاحتمال هذا الخطأ بالرمز  $\beta$  .

وقد اصطلح على تسمية المقدار  $1 - \beta$  بقوة الاختبار power of the test وهو يعبر عن احتمال تجنب الخطأ من النوع الثاني . ومن الواضح أن قوة الاختبار تزيد كلما نقص الاحتمال  $\beta$  للخطأ من النوع الثاني والعكس بالعكس .

نلخص المعاني السابقة في الجدول (٧ - ١) الآتي :

الجدول (٧ - ١)

حالات رفض أو قبول الفرض العفري

الحالة المجهولة للفرض العفري	القرار
ف. صحيح      ف. زائف	
خطأ I (باحتمال $\alpha$ )      صواب (باحتمال $1 - \beta$ )	رفض ف.
صواب (باحتمال $1 - \alpha$ )      خطأ II (باحتمال $\beta$ )	قبول ف.

يلاحظ أن كلا من الاحتمالين  $\alpha$  ،  $\beta$  هو احتمال شرطى ونكتب :

$\alpha = \text{ل ( رفض ف. | ف. صحيح )}$  ،  $\beta = \text{ل ( قبول ف. | ف. زائف )}$  .

في الأمثلة التي تناولناها في الفصل السابق كان اهتمامنا منصبا على الخطأ من النوع الأول وحرصنا على أن يكون الاحتمال  $\alpha$  لهذا الخطأ عددا صغيرا يسمح لنا بالتجاوز عن هذا الخطأ ، وسمينا هذا العدد مستوى الدلالة واتخذناه كأحد أسس قاعدة اختبار الفروض ، خاصة فيما يتعلق بفصل فضاء العينة إلى منطقتي رفض وقبول الفرض العفري . وسنهم الآن بالخطأ من النوع الثاني ، إذ ينبغي عند تصميم التجارب وبناء اختبارات الفروض أن نعمل على أن يكون كل من الاحتمالين  $\alpha$  ،  $\beta$  صغيرا على قدر الإمكان .

غير أنه نظرا لأن تصغير أحد هذين الاحتمالين يؤدي إلى كبر الآخر كما سنرى في البند (٧ - ٢ - ٢) ، يكون من العبث البحث عن طريقة عامة تضمن صغر كل من هذين الاحتمالين معا ونكون حينئذ أمام مشكلة يجب أن نجد لها حلا .

والطريقة المعتادة لتناول هذه المشكلة تبدأ بالتحكم في احتمال الخطأ من النوع الأول ( وهو الخطأ الأكثر خطورة ) وذلك بوضع حد أعلى للاحتمال  $\alpha$  فنختار لهذا الحد قيمة صغيرة مثل ٠,٠٥ أو ٠,٠١ ثم نحسب على أساسها كلا من منطقتي قبول ورفض الفرض الصفري من واقع ما لدينا من بيانات ومن معرفتنا بتوزيع إحصاء الاختبار . بعد ذلك نقوم بحساب الاحتمال  $\beta$  ، فإذا كان هذا الاحتمال صغيرا تكون المشكلة قد حلت تلقائيا ، أما إذا كان كبيرا بدرجة لا نستطيع معها المجازفة به وجب علينا أن نعمل على تخفيض هذا الاحتمال . وهذا ما سنتناوله في البند ( ٧ - ٥ ) بعد تقديم طريقة حساب هذا الاحتمال والعوامل المؤثرة عليه .

### طريقة إيجاد احتمال الخطأ من النوع الثاني :

نفرض أننا اخترنا مستوى الدلالة  $\alpha$  وحددنا عند هذا المستوى كلا من منطقة الرفض  $^2$  ومنطقة القبول  $^2$  للفرض الصفري مع ملاحظة أن بارامتر إحصاء الاختبار يتحدد هنا على أساس التسليم بصحة هذا الفرض . بعد ذلك نوجد احتمال وقوع قيم المتغير في منطقة القبول  $^2$  على أساس أن الفرض الصفري زائف وأن الفرض الآخر هو الصحيح فيكون هذا الاحتمال هو احتمال قبول الفرض الصفري عندما يكون زائفا أى هو الاحتمال  $\beta$  للخطأ من النوع الثاني مع ملاحظة اختلاف بارامتر إحصاء الاختبار . وعلى هذا فحساب الاحتمال  $\beta$  يكون على خطوتين هما :

( أ ) تحديد منطقة القبول على أساس صحة الفرض الصفري ،

( ب ) إيجاد احتمال وقوع المتغير في هذه المنطقة على أساس أن الفرض الآخر هو الصحيح .

سنوضح هذا الأسلوب لحساب قيمة الاحتمال  $\beta$  في عدة حالات نبدأها في البند ( ٧ - ٢ ) بحالة الاختبار ذي الجانب الواحد لفرض عن متوسط مجتمع معتدل

مع بيان العوامل المؤثرة على قوة الاختبار وكيفية زيادة هذه القوة ، ثم نطبق ذلك كله على ثلاث حالات أخرى نقدمها في البنود الثلاثة الأخيرة .

(٧ - ٢) حساب قيمة  $\beta$  في اختبار فرض عن متوسط معتدل - حالة الاختبار ذو الجانب الواحد .

مثال (٧ - ١) :

اعتبر مجتمعا معتدلا تباينه  $\sigma^2 = ٤٩$  ووسطه الحسابى  $\mu$  غير معروف . بين كيف تستخدم الوسط الحسابى  $\bar{x}$  لعينة عشوائية من الحجم  $n = ٢٥$  لاختبار الفرض الصفري  $\mu = ٧٥$  ضد الفرض الآخر  $\mu < ٧٥$  عند مستوى الدلالة  $\alpha = ٥\%$  ثم أوجد قوة الاختبار عندما  $\mu = ٧٦$  .

الحل :

نظرا لأن المجتمع معتدل فإن المتغير  $\bar{x}$  الذى يعبر عن الوسط الحسابى للعينات من الحجم  $n$  يكون ذا توزيع معتدل وسطه الحسابى  $\mu$  وانحرافه المعياري  $\frac{\sigma}{\sqrt{n}}$  - راجع البند (٦ - ٣) - وبالتالي يكون للإحصاءة

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

توزيع معتدل معيارى : مع (٠ ، ١) .

لدينا  $\sigma^2 = ٤٩$  ،  $n = ٢٥$  ،  $\alpha = ٠,٠٥$  وإذن  $\frac{\sigma}{\sqrt{n}} = \frac{٧}{\sqrt{٢٥}} = ١,٤$

الفرض الصفري  $F_0 : \mu = ٧٥$

الفرض الآخر  $F_1 : \mu < ٧٥$

نظرا لأن الاختبار ذو جانب واحد هو الجانب الأيمن فإن منطقة الرفض  $R$  هي

تلك المنطقة التي يأخذ فيها الوسط الحسابي  $\bar{x}$  لعينة من الحجم ٢٥ قيمة أكبر من العدد ١ حيث :

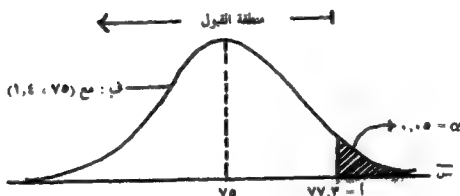
$$L (\bar{x} < 1) = 0,05 \quad \text{تحت الفرض الصفري } \mu = 70$$

$$\text{ومنها } L = \left( \frac{70 - 1}{1,4} < \frac{70 - \bar{x}}{1,4} \right) = 0,05$$

$$\text{أى } L (E) = \left( \frac{70 - 1}{1,4} < \right) = 0,05$$

من جدول المساحات أسفل المنحنى المعتدل المعيارى نجد أن

$$\frac{70 - 1}{1,4} = 1,64 = \frac{70 - \bar{x}}{1,4} \quad \text{ومنها } 1 = 77,296 = 77,3 \text{ تقريبا}$$



الشكل (٧-١)

منطقة القبول والرفض في اختبار ذي جانب واحد

وإذن المنطقة التي نرفض فيها الفرض الصفري  $\mu = 70$  حين يكون هذا الفرض صحيحا وعند المستوى  $\alpha = 0,05$  هي تلك المنطقة التي تأخذ فيها الأوساط الحسابية للعينات ذوات الحجم ٢٥ قيمة تزيد عن ٧٧,٣ واحتمال وقوع المتغير

في هذه المنطقة هو ٥٪، وتعبّر عن هذا الاحتمال مساحة الجزء المظلل بالشكل (١ - ٧) . وعلى ذلك تتحدد قاعدة الاختبار كالآتي :

« إذا وقع الوسط الحسابي  $\bar{x}$  لعينة عشوائية من الحجم ٢٥ في المنطقة ٢ =  $\{ \bar{x} : \bar{x} < ٧٧,٣ \}$  نرفض الفرض الصفري  $H_0$  عند مستوى الدلالة ٥٪ ولا تقبل  $H_0$  » .

أما منطقة قبول الفرض الصفري فهي بالطبع المنطقة  $\bar{x} \geq ٧٧,٣$  المكتملة لمنطقة الرفض . أى أننا نقبل الفرض الصفري إذا كان الوسط الحسابي لعينة عشوائية من الحجم ٢٥ مأخوذة من التوزيع المثل لهذا الفرض يساوى أو يقل عن ٧٧,٣ ونسبة هذه العينات هي ٩٥٪ من العينات التي تؤخذ من هذا التوزيع .

ولكن ماذا لو كان الفرض الصفري راثفا والفرض الآخر هو الصحيح ؟ نرفض مثلا أن القيمة الصحيحة هي  $\mu = ٧٦$  ( وهذه قيمة تحقق الفرض الآخر  $\mu < ٧٥$  ) .

في هذه الحالة يكون للمتغير  $\bar{x}$  توزيع معتدل : مع (١,٤ ، ٧٦) ، ومنتج ما يلي :

$\beta$  = احتمال قبول الفرض الصفري عندما يكون راثفا والفرض الآخر

$\mu = ٧٦$  هو الصحيح

= احتمال وقوع قيم المتغير  $\bar{x}$  في منطقة القبول  $\bar{x} \geq ٧٦$  حين يكون لهذا المتغير

توزيع معتدل مع (١,٤ ، ٧٦)

$J = P(\bar{x} \geq ٧٧,٣)$  حيث  $\bar{x} \sim (١,٤ ، ٧٦)$

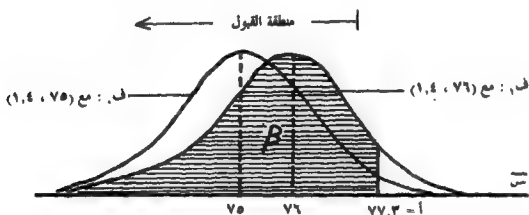
$$J = P(\bar{x} \geq ٧٧,٣) = P\left(\frac{\bar{x} - ٧٦}{١,٤} \geq \frac{٧٧,٣ - ٧٦}{١,٤}\right) = P(Z \geq ٠,٩٣) = ٠,٨٢٣٨$$

من جدول المساحات أسفل المنحنى المعتدل المعيارى نجد أن

$$\beta = ٠,٨٢٣٨ = ٠,٨٢ \text{ تقريبا}$$

وهذا هو احتمال الوقوع في الخطأ من النوع الثاني . أما قوة الاختبار عندما تأخذ  $\mu$  القيمة ٧٦ فهي  $1 - \beta = 0.18$

لتوضيح ذلك هندسيا نرسم كما في الشكل (٧ - ٢) توزيع المتغير  $\bar{x}$  في حالتين ، أولاهما عندما يكون الفرض الصفري ف صحيحا ويكون التوزيع معتدلا متوسطه ٧٥ وثانيهما عندما يكون الفرض الآخر ف صحيحا ويكون التوزيع معتدلا متوسطه ٧٦ .



الشكل (٧ - ٢)

التوزيع المعتدل للفرض الصفري والتوزيع المعتدل للفرض الآخر  
(مساحة الجزء المظلل صير عن الاحتمال  $\beta$  في اختبار ذي جانب واحد)

من هذا الشكل يتبين أن بعض العينات التي تنتمي إلى توزيع ف٠ تكون متوسطاتها واقعة في منطقة القبول لتوزيع ف٠ ونسبة هذه العينات هي نسبة الجزء من توزيع ف٠ الذي يشترك مع توزيع ف٠ في منطقة القبول ، وهي تعطى بالاحتمال  $L$  ( $L \geq 0.77,3$ ) محسوبا من توزيع ف٠ وهذا الاحتمال هو بالضبط احتمال قبول الفرض الصفري عندما يكون الفرض الآخر هو الصحيح ، أي هو الاحتمال  $\beta$  للخطأ من النوع الثاني .



( يلاحظ أننا لا نستطيع حساب قيمة  $\beta$  إلا إذا حددت قيمة معينة مثل  $\mu = 76$  للبارامتر  $\mu$  تحقق الفرض الآخر ( وهو  $\mu < 75$  ) وذلك لكي يتحدد التوزيع الممثل للفرض الآخر تحديدا تاما .

## ( ٧ - ٢ - ١ ) دالة القوة POWER FUNCTION

في هذا المثال وجدنا أن قوة الاختبار عندما نفترض أن  $\mu = 76$  هي ٠,١٨ . وتختلف هذه القوة بحسب قيمة  $\mu$  فإذا افترضنا أن  $\mu = 77$  نجد بنفس المنطق السابق أن

$$\beta = J(\bar{S} \geq 77,3) \quad \text{حيث } \bar{S} : \text{ مع } (77, 1,4)$$

$$J = \frac{77 - \bar{S}}{1,4} \geq \frac{77 - 77,3}{1,4} = J(8 \geq 0,21)$$

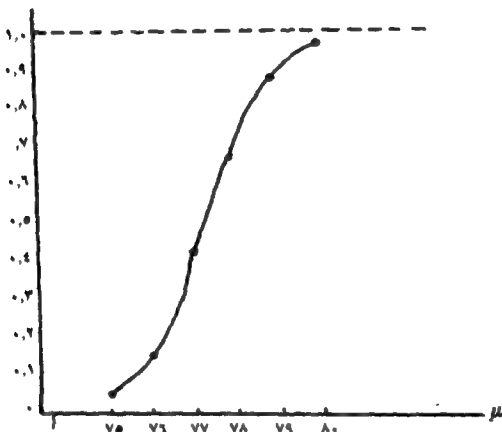
$= 0,58$  وبالتالي يكون قوة الاختبار  $1 - 0,58 = 0,42$  وبالمثل : بأخذ  $\mu = 78$  نجد أن  $\beta = 0,31$  و  $1 - \beta = 0,69$  وبأخذ  $\mu = 79$  نجد أن  $\beta = 0,11$  و  $1 - \beta = 0,89$  وبأخذ  $\mu = 80$  نجد أن  $\beta = 0,03$  و  $1 - \beta = 0,97$  ... إلخ ومن هذا نرى أن قوة الاختبار تتوقف على البارامتر  $\mu$  أى هي دالة د في  $\mu$  تأخذ الصيغة الآتية :

$$\text{دالة قوة الاختبار} = د(\mu) = 1 - \beta = 1 - J(\bar{S} \geq 1) \\ J(\bar{S} < 1) = \text{حيث } \bar{S} : \text{ مع } (\mu, 1,4)$$

وحيث  $1$  هي القيمة الحرجة التى تفصل بين منطقتي القبول والرفض . ( قيمة الدالة د عند قيمة معينة  $\mu$  تسمى قوة الاختبار عند القيمة  $\mu$  ) .

وتمثل دالة القوة بيانياً كما في الشكل (٧ - ٣) الذي يوضح أنها دالة تزايدية ،  
تزيد قيمتها كلما بعدت قيمة  $\mu$  التي يحددها الفرض الآخر عن قيمة  $\mu$  التي يحددها  
الفرض الصفري .

د (  $\mu$  ) = قوة الاختبار



الشكل (٧ - ٣)

منحنى القوة لاختبار المثال (٧ - ١) - اختبار ذو جانب واحد

في هذا المثال وجدنا أن احتمال الخطأ من النوع الثاني  $\beta = 0.82$  ، ومن الواضح  
أن هذا الاحتمال هو احتمال كبير لهذا الخطأ لا ينبغي أن نسمح به حين نتخذ قراراً

بشأن رفض أو قبول الفرض الصفري لأنه يجعلنا نشك في حساسية الاختبار ، بمعنى أنه إذا كانت  $\alpha = 0.05$  ،  $\sigma = 49$  وأخذنا عينة من الحجم  $n = 25$  فإن هذه العينة لا تكون قادرة على التمييز بين الفرضين بدرجة كافية من الثقة ، إذ بالرغم من أن  $95\%$  من العينات المأخوذة من التوزيع الذى يفترض صحة الفرض الصفري ( $\mu = 75$ ) تقع في منطقة القبول ، إلا أن  $82\%$  من العينات المأخوذة من التوزيع الذى يفترض صحة الفرض الآخر ( $\mu = 76$ ) تقع أيضا في المنطقة ذاتها . وهذا التداخل الكبير هو الذى نعينه بقولنا إن الاختبار ذو قوة ضعيفة أو أنه اختبار غير حساس .

في مثل هذه الحالة يجب أن ندخل تعديلا في تصميم التجربة التى تمدنا بالبيانات التى نتخذ قرارنا على أساسها لتلافي الوقوع في خطأ كبير من النوع الثانى ولإعطاء الاختبار قوة كافية للتمييز بين مختلف الفروض . وفى بحثنا عن التعديل اللازم لتحقيق هذا الفرض نبدأ بتدريس العوامل التى تؤثر في هذا الخطأ .

## (٧ - ٢ - ٢) العوامل المؤثرة في الخطأ من النوع الثانى :

بالتأمل في المثال السابق يتبين لنا أن الاحتمال  $\beta$  للخطأ من النوع الثانى وقوة الاختبار  $\phi$  يتوقفان على القيم الآتية :

### (١) القيمة التى تختار للاحتمال $\alpha$ للخطأ من النوع الأول :

ذلك لأن هذه القيمة هى التى تحدد القيمة الحرجة  $t$  ( تساوى  $77.3$  في هذا المثال ) التى تفصل بين منطقتى الرفض والقبول . وكلما صغرت قيمة  $\alpha$  كلما صغرت منطقة الرفض وأزيمت  $t$  إلى اليمين ( في هذا المثال ) واتسعت منطقة القبول وبالتالي زاد احتمال هذه المنطقة تحت الفرض الآخر . ومن هنا يمكننا تقبل الحقيقة الآتية :

كلما صغر الاحتمال  $\alpha$  للخطأ من النوع الأول كلما كبر الاحتمال  $\beta$  للخطأ من النوع الثاني وصغرت قوة الاختبار .

وهذه الحقيقة تعطينا سببا وجها يبرر التقليد المتبع باختيار  $\alpha = 0,05$  أو  $\alpha = 0,01$  ، لأن اختيارنا لقيم تقل عن ذلك يؤثر تأثيرا سيفا في الاحتمال  $\beta$  الذى يمكن التحكم فيه بتكبير حجم العينة كما سيتبين بعد .

### (٢) قيمة البارامتر $\mu$

بالتأمل في قيم  $\beta$  أو  $\alpha$  التى حسبناها في البند (٧ - ٢ - ١) نجد أن هذه القيم تتوقف على بعد القيمة  $\mu$  التى يحددها الفرض الآخر عن القيمة  $\mu_0$  التى يحددها الفرض الصفري . ومن الناحية الهندسية إذا كانت  $\mu_0$  ،  $\mu$  قريتين من بعضهما أى كان متوسطا توزيعى  $F_1$  ،  $F_2$  قريين من بعضهما فإن التداخل بين هذين التوزيعين في منطقة القبول يكون كبيرا وهذا يؤدي إلى كبر الاحتمال  $\beta$  وصغر قوة الاختبار . أما إذا كانت  $\mu$  بعيدة عن  $\mu_0$  فإن هذا التداخل يكون صغيرا ويؤدي إلى صغر الاحتمال  $\beta$  وكبر قوة الاختبار . وتتضح هذه الحقيقة أيضا عند التأمل في منحنى دالة القوة المبين بالشكل (٧ - ٣) . ومن هنا يمكننا تقبل الحقيقة الآتية :

كلما زاد الفرق بين القيمة التى يحددها الفرض الصفري للبارامتر المختبر والقيمة التى يحددها الفرض الآخر ، كلما صغر الاحتمال  $\beta$  وزادت قوة الاختبار

### (٣) الخطأ المعياري لإحصاءة الاختبار :

في المثال (٧ - ١) كان الخطأ المعياري لإحصاءة الاختبار هو  $\sigma_{\bar{y}} = 1,4$  ووجدنا أن قيمة العدد  $z$  الذى يفصل بين منطقتى الرفض والقبول هي  $z = 1,77$  . إذا

أجرينا تعديلا في هذا المثال بحيث يصبح الخطأ المعياري أقل من ١,٤ نجد أن  $\sigma$  تصغر وتزاح النقطة المثلثة لها على توزيع الإحصاء  $\bar{X}$  إلى اليسار وبالتالي تصغر منطقة القبول ويقل الاحتمال  $\beta$  . فمثلا إذا أخذنا  $\sigma = 1,1$  فإن

$$L(\bar{X} < 1) = 0,05 \quad \text{ومنها} \quad L(E < 1) = \left( \frac{1 - 0,05}{1,1} \right) = 0,05$$

$$\therefore \frac{1 - 0,05}{1,1} = 1,64 \therefore 1 = 1,64 \times 1,1 = 1,804 \text{ وهذا العدد أصغر من } 1,804 \text{ ومن}$$

هنا يمكننا تقبل الحقيقة الآتية :

كلما صغر الخطأ المعياري لإحصاء الاختبار  
كلما صغر الاحتمال  $\beta$  وزادت قوة الاختبار .

ومن الواضح أن قيمة الكسر  $\frac{\sigma}{\sqrt{n}}$  تتوقف على قيمتين هما الانحراف المعياري  $\sigma$  للمجتمع وحجم العينة  $n$  ، ويصغر هذا الكسر ( أى يصغر الخطأ المعياري ) إذا صغرت قيمة  $\sigma$  فقط أو كبرت قيمة  $n$  فقط أو صغرت  $\sigma$  وكبرت  $n$  في الوقت نفسه .

### (٧ - ٢ - ٣) كيفية زيادة قوة الاختبار :

في الفقرة السابقة وجدنا أن قوة الاختبار تتوقف على أربع قيم هي  $\mu$  ،  $\sigma$  ،  $n$  . ولما كانت القيمتان  $\mu$  ،  $\sigma$  تحدّدان سلفا بحسب خبرة الباحث وطبيعة المشكلة التي يتناولها ، لا يبقى لدينا من التاحية الإحصائية لزيادة قوة الاختبار إلا الاعتماد على تصغير الخطأ المعياري  $\frac{\sigma}{\sqrt{n}}$  وذلك بتصغير  $\sigma$  أو تكبير  $n$  .

ونظرا لأننا نقدر عادة تبين المجتمع  $\sigma^2$  من تبين العينة فإن زيادة قوة الاختبار تقتضى أن نحرص على ألا يكون هذا التقدير أكبر مما ينبغي وهذا لا يتأتى إلا

بالتحكم الجيد في ظروف التجريب واستبعاد تأثير أية عوامل خارجية تؤثر في المشاهدات وتسهم في زيادة تباينها .

أما زيادة حجم العينة فهو العامل الرئيسى الذى نعتد عليه في زيادة قوة الاختبار ، وهذه أهم نتيجة نخرج بها من هذا الفصل وتتلخص في الحقيقة الآتية :

إذا تساوت جميع الظروف ، كلما كبر حجم العينة  
كلما صغر الاحتمال  $\beta$  وزادت قوة الاختبار .

### الحد الأمثل لحجم العينة :

على أن كفاءة التجريب تستدعى ألا يكون حجم العينة أكبر مما ينبغي تحسباً لما تتطلبه هذه العملية من جهد ووقت وتكاليف ، ومن المناسب إذن وضع حد أعلى لحجم العينة يحقق المهدف المنشود من زيادة قوة الاختبار دون تحمل أعباء لا ضرورة لها . غير أنه لا توجد قاعدة عامة لتحديد الحجم المناسب للعينة ، إلا أننا نستطيع ذلك في بعض الحالات الخاصة ومنها الحالات التى يتناولها هذا الفصل .

ففى حالة اختبار فرض عن متوسط مجتمع معتدل ، نفرض أننا حددنا مسبقاً قيمة  $\alpha$  وقيمتى الفرض الصفري والفرض الآخر . إذا أردنا أن نضمن أن يأخذ احتمال الخطأ من النوع الثانى قيمة محددة  $\beta$  يمكن إثبات أن الحد الأعلى لحجم العينة الذى يحقق هذا الضمان ينتج من حل المعادلة الآتية :

$$(1) \quad \frac{f}{x_1 - x_2} = 0.2 \cdot x$$

حيث  $x = 2$  = الخطأ المعياري لإحصاء الاختبار  
،  $f$  = الفرق بين القيمة التى يحددها الفرض الصفري والقيمة التى يحددها الفرض الآخر

أما  $\bar{E}$  ،  $E$  ، فتحددان محل المتباينتين الآتيتين :

$$L = (E < \bar{E}) = \left. \begin{array}{l} \alpha \text{ إذا كان الاختبار ذا جانب واحد} \\ \frac{\alpha}{2} \text{ إذا كان الاختبار ذا جانبيين} \end{array} \right\}$$

$$L = (E \geq \bar{E}) = \beta$$

حيث  $\bar{E}$  هو المتغير المعتدل المعياري : مع (٠ ، ١) . أى أن  $\bar{E}$  ،  $E$  ، توجدان من جدول المساحات أسفل المنحنى المعتدل المعياري بمجرد التعويض عن قيمتي  $\alpha$  ،  $\beta$  . والمعادلة (١) هى معادلة عامة فى حالة اختبار فرض عن متوسط مجتمع معتدل أو اختبار الفرق بين متوسطى مجتمعين معتدلين ، مع بعض الاختلافات فى حساب القيم التى تتركب منها هذه المعادلة كما سنرى بعد .

مثال (٧ - ٢) :

اعتبر المثال (٧ - ١) حيث  $\sigma^2 = ٤٩$  ،  $\alpha = ٠,٠٥$  نريد اختبار الفرض  $\mu = ٧٥$  ضد الفرض  $\mu < ٧٥$  . وإذا أخذنا  $\mu = ٧٨$  فأوجد الحد الأعلى لحجم العينة الذى يضمن أن تكون  $\beta = ٠,٠٨$  .

الحل :

$$X = \frac{\bar{Y}}{\sigma_Y} = ٠,٢$$

$$F = \mu - \bar{\mu} = ٧٥ - ٧٨ = ٣$$

من جدول المساحات أسفل المنحنى المعتدل المعياري نجد أن :

$L = (E < \bar{E}) = \alpha = ٠,٠٥$  تعطى  $\bar{E} = ١,٦٤$  ( اختبار ذو جانب واحد )

$$L = (E \geq \bar{E}) = \beta = ٠,٠٨ \text{ تعطى } \bar{E} = ١,٤١$$

بالتعويض في المعادلة (١) ينتج أن :

$$\frac{3}{3,00} = \frac{3}{(1,41) - 1,64} = \frac{7}{\sqrt{7}}$$

$$0,6469 = \frac{(3,00 \times 7)}{3} = 7 \therefore$$

أى أننا إذا أعدنا التجربة بأخذ عينة عشوائية من الحجم ٥١ (بدلاً من الحجم ٢٥) فإننا نضمن أن يكون احتمال الخطأ من النوع الثانى  $\beta = 0,08$  (تحقق من ذلك بأخذ  $h = 0,51$  وإثبات أن القيمة الحرجة  $z = 1,64$  والاحتمال  $\beta = 0,0778$ ). يلاحظ أننا استطعنا أن نضمن احتمالاً صغيراً هو  $0,08$  للخطأ من النوع الثانى (بدلاً من الاحتمال  $\beta = 0,31$  الذى وجدناه في عينة بالحجم ٢٥)، إلا أن ذلك كان على حساب زيادة حجم العينة إلى الضعف تقريباً في هذا المثال.

(٧ - ٣) حساب قيمة  $\beta$  في اختبار فرض عن متوسط مجتمع معتدل - حالة الاختبار ذى الجانبين .  
مثال (٧ - ٣) :

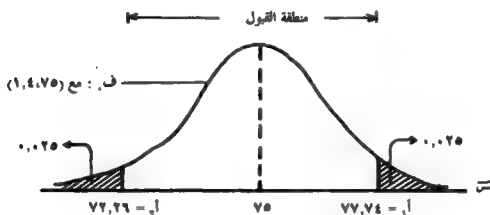
بأخذ بيانات المثال (٧ - ١) يبين كيف تستخدم الوسط الحسابى لعينة لاختبار الفرض الصفري ف :  $\mu = 70$  ضد الفرض الآخر  $\mu \neq 70$  عند مستوى الدلالة  $0,05$  . حدد قوة الاختبار عند  $\mu = 76$  وارسم دالة القوة لهذا الاختبار .

الحل :

نظراً لأن المجتمع معتدل يكون للمتغير  $\bar{x}$  توزيع معتدل : مع  $(\frac{\sigma}{\sqrt{n}}, \mu)$  وبالتالي يكون للاحصاءة 
$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$
 توزيع معتدل معيارى : مع  $(0, 1)$  .



ونظرا لأن الاختبار ذو جانبيين فإن منطقة رفض الفرض الصغرى تتألف من جزعين متساويين في جانبي التوزيع . لتكن  $\alpha$  هي القيمة الحرجة التي تحدد منطقة الرفض اليمنى من اليسار ، ولتكن  $\alpha$  هي القيمة الحرجة التي تحدد منطقة الرفض اليسرى من اليمين . انظر الشكل (٧ - ٤) .



الشكل (٧ - ٤)

منطقتا الرفض والقبول في اختبار ذي جانبيين

لإيجاد قيمتي  $\alpha$  ،  $\alpha$  نستخدم جدول المساحات أسفل المنحنى المعتدل المعياري كالتالي :

$$\text{ل } (\bar{x} < \alpha) = \frac{70 - \alpha}{1,4} = 0,025 \quad \text{حيث } \bar{x} = (75, 1,4)$$

$$\therefore \text{ل } \left( \frac{70 - \alpha}{1,4} < \frac{70 - \bar{x}}{1,4} \right) = 0,025 \quad \text{أي ل } \left( \frac{70 - \alpha}{1,4} < \frac{70 - 75}{1,4} \right) = 0,025$$

$$\therefore \frac{70 - \alpha}{1,4} = 1,96 \quad \text{ومنها } \alpha = 77,744$$

بالمثل ، ل (  $\bar{S} > 1$  )  $= \frac{0.05}{3} = 0.025$  حيث  $\bar{S}$  : مع ( ٧٥ ، ١,٤ )

$$\therefore \text{ل ( } \bar{E} > 1,4 = \frac{75 - 1}{1,4} > 0.025$$

$$\therefore 1,96 - \frac{75 - 1}{1,4} = 72,206 = 1,4$$
 ومنها

وإذن تتحدد منطقة قبول الفرض الصفري حين يفترض صحته بالفترة ( ٧٢,٢٦ ، ٧٧,٧٤ ) وتكون قاعدة الاختبار كالآتي :

• إذا وقع الوسط الحسابي لعينة عشوائية من الحجم ٢٥ خارج المنطقة  $\bar{S}' = \{ \bar{S} : 77,74 \leq \bar{S} < 72,26 \}$  نرفض الفرض الصفري ف عند مستوى الدلالة ٠,٠٥ ، وإلا نقبل ف ،

نحسب الآن الاحتمال  $\beta$  عند  $\mu = 76$  وفقا للمنطق السابق .  
 $\beta$  = احتمال وقوع قيم المتغير  $\bar{S}$  في منطقة القبول ( ٧٢,٢٦ ، ٧٧,٧٤ ) تحت الفرض الآخر  $\mu = 76$

$$= \text{ل ( } 77,74 < \bar{S} < 72,26 \text{ ) حيث } \bar{S} : \text{ مع ( } 76 , 1,4 \text{ )}$$

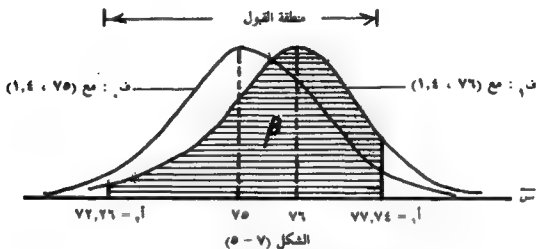
$$= \text{ل ( } \frac{76 - 77,74}{1,4} \leq \frac{\bar{S} - 76}{1,4} \leq \frac{76 - 72,26}{1,4} \text{ )}$$

$$= \text{ل ( } 1,24 \leq \bar{E} \leq 2,67 \text{ )} = 0,89 \text{ تقريبا}$$

$$\therefore \text{قوة الاختبار عند } \mu = 76 \text{ هي } 1 - 0,89 = 0,11 \text{ تقريبا}$$

لتوضيح ذلك هندسيا نرسم كما في الشكل (٧ - ٥) توزيع المتغير  $\bar{S}$  تحت كل من الفرضين :  $\mu = 75$  وف :  $\mu = 76$  من هذا الشكل يلاحظ

أن الاحتمال  $\beta = 0.89$  هو نسبة الجزء من توزيع  $F_1$  الذى يشترك في منطقة القبول مع توزيع  $F_2$  ، وهذه النسبة تعبر عنها مساحة الجزء المظلل من الشكل (٥ - ٧) .



التوزيع المظلل للفرض الصفري والتوزيع المظلل للفرض الآخر  
(مساحة الجزء المظلل تعبر عن الاحتمال  $\beta$  في اختبار  $t$  جانين)

في هذه الحالة تأخذ دالة قوة الاختبار الصيغة الآتية :

$$D(\mu) = 1 - \Phi\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right) + \Phi\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)$$

$$= 1 - \Phi\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right) + \Phi\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)$$

وفي هذا المثال نجد أن :

$$D(\mu) = 1 - \Phi\left(\frac{77.74 - 75}{1.4/\sqrt{10}}\right) + \Phi\left(\frac{72.26 - 75}{1.4/\sqrt{10}}\right) \quad \text{حيث } \bar{X} = (1.4, \mu)$$

$$\text{وقد وجدنا أن } D(76) = 0.89 - 1 = 0.11$$

وبالمثل نجد ما يلي :

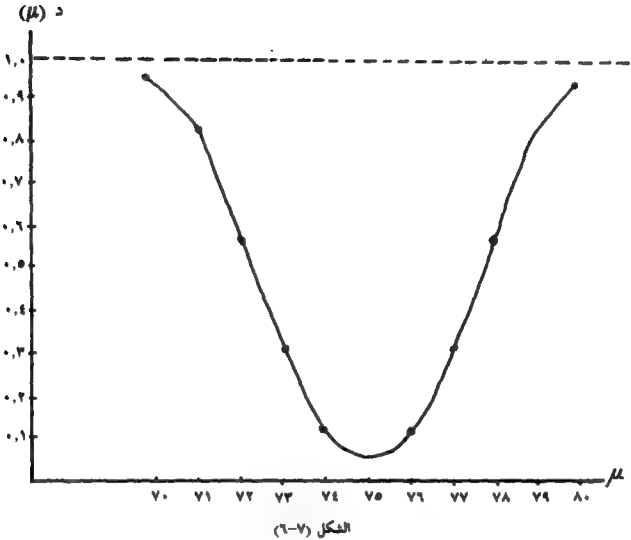
$$D(77) = 0.30 \text{ و } D(78) = 0.07 \text{ و } D(79) = 0.82 \text{ و } D(80) = 0.94$$

... الخ . ومن الملاحظ نجد أن

$$D(74) = 0.11 \text{ و } D(73) = 0.30 \text{ و } D(72) = 0.07 \text{ و } D(71) = 0.82$$

$$\text{و } D(70) = 0.94 \dots \text{ الخ}$$

والشكل (٧ - ٦) يمثل هذه الدالة بيانيا .



مسحى القوة لاختبار الخال (٧-٧) - اختبار ذو جانين

إن الملاحظات والحقائق التي ذكرت في البند (٧ - ٢ - ٢) عن العوامل المؤثرة في قيمة  $\beta$  تنطبق هنا أيضا . فكلما صغر الاحتمال  $\alpha$  كلما صغر جزءا منطقة الرفض ، واتسعت منطقة القبول وهذا يؤدي إلى زيادة الاحتمال  $\beta$  وصغر قوة الاختبار . كذلك كلما بعدت القيمة  $\mu$  التي يفرضها الفرض الآخر عن القيمة التي يحددها الفرض الصفرى كلما قل التداخل بين توزيعي ف ، ف ، وصغرت

$\beta$  وزادت قوة الاختبار . وأخيرا كلما تقلص الخطأ المعياري سواء بتصغير الانحراف المعياري  $\sigma$  للمجتمع أو بزيادة حجم العينة ، كلما صغرت  $\beta$  وزادت قوة الاختبار .

وجدير بالملاحظة هنا أنه إذا تساوت جميع الظروف فإن الاحتمال  $\beta$  للخطأ من النوع الثاني يكون أقل في الاختبار ذي الجانب الواحد منه في الاختبار ذي الجانبين ، أى أن الاختبار ذا الجانب الواحد يكون أقل تعرضا لهذا النوع من الخطأ .

وكما في البند (٧ - ٢ - ٣) حين نتناول اختبارا ذا جانبيين لفرض عن متوسط مجتمع معتدل ، إذا تحددت قيم  $\alpha$  ،  $\mu$  ،  $\sigma$  وأردنا أن نضمن أن يأخذ احتمال الخطأ من النوع الثاني قيمة محددة  $\beta$  فإن الحد الأعلى لحجم العينة الذى يوفر هذا الضمان هو ذلك الذى ينتج من حل نفس المعادلة (١) السابق تقديمها وهى :

$$\frac{f}{\bar{c}_1 - \bar{c}_2} = 0.2$$

والفرق في استخدام هذه المعادلة بين الحالتين يحدث فقط في حساب القيمة  $\bar{c}_1$

مثال (٧ - ٤) :

في المثال (٧ - ٣) حيث  $\sigma = 49$  ،  $\alpha = 0.05$  ،  $\mu = 70$  إذا أخذنا  $\mu = 78$  فأوجد الحد الأعلى لحجم العينة الذى يضمن أن تكون  $\beta = 0.10$  .

الحل :

$$\frac{y}{\sigma_y} = \frac{g}{\sigma_y} = 0.2 \text{ خ}$$

$$f = \mu - \mu = 78 - 70 = 8$$

$$\begin{aligned} L(ع < ع) &= \frac{0.40}{1} = 0.40 \text{ ومنها } 1.96 = \text{ (اختبار ذو جانبي)} \\ L(ع \geq ع) &= \frac{0.10}{1} = 0.10 \text{ ومنها } 1.28 = \end{aligned}$$

$$\therefore \frac{1}{1.08} = \frac{2}{3.24} = \frac{3}{(1.28) - 1.96} = \frac{7}{0.68}$$

ومنها  $0.10 = (1.08 \times 7) = 0.756$   
 أى أنه يمكن أن نأخذ عينة حجمها ٥٨ لنضمن أن تكون  $\beta = 0.10$  (تحقق  
 من ذلك بأخذ  $0.08$  وإثبات أن القيمتين الخارجيتين هما  $0.7680$  ،  
 $0.981 = \beta$  وأن  $0.73199$  )

(٧ - ٤) حساب قيمة  $\beta$  في اختبار الفرق بين متوسطي مجتمعين  
 مختلفين :

نستخدم نفس الأسلوب المقدم في البتين السابقين مع مراعاة طريقة حساب  
 الخطأ المعياري التي تتطلبها هذه الحالة .

مثال (٧ - ٥) :

لتجربة أقراص لإنقاص الوزن عند النساء ، اختبرت ٤٠ من إناث القطة المنزلية  
 وقسمت عشوائيا إلى مجموعتين بكل منها ٢٠ قطة ووضعت المجموعتان تحت نفس  
 الظروف والنظام الغذائي فيما عدا أن الأقراص كانت تضاف إلى غذاء واحدة فقط  
 من المجموعتين . وبعد ستة أسابيع وباستخدام مقياس معين حسب الوسيطان  
 الحسابيان  $\bar{X}_1$  ،  $\bar{X}_2$  للمجموعتين . إذا اعتبرنا أن المجموعتين مستقلتان  
 وأخوذتان من مجتمعين مختلفين تبين كل منها ٤٠ ومتوسطاهما  $\mu_1$  ،  $\mu_2$  :  
 (أولا) بين كيف نختبر الفرض الصفري  $\mu_1 = \mu_2$  ضد الفرض  $\mu_1 \neq \mu_2$  :  
 عن مستوى الدلالة ٠.٠٥ .

( ثانيا ) أوجد قوة الاختبار عندما يفترض أن الفرق بين متوسطي المجتمعين يساوى ٣ وحدات .

( ثالثا ) أوجد الحد الأدنى لحجم العينة الذى يضمن أن يكون الاحتمال  $\beta$  للخطأ من النوع الثانى يساوى ٠,١٥ .

الحل :

( أولا )

نظرا لأن المجتمعين معتدلان والعيتين مستقلتان فإن المتغير العشوائى  $\bar{y}_1 - \bar{y}_2$  يكون ذا توزيع معتدل وسطه الحسابى  $\mu_1 - \mu_2$  وتباينه يساوى

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \quad \text{لأن } \sigma^2 = \sigma_1^2 = \sigma_2^2 \quad \text{و } \sigma = \sigma_1 = \sigma_2$$

وبالتالى يكون للإحصاءة

$$\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\frac{\sigma^2}{n}} = z$$

توزيع معتدل معيارى : مع ( ١ ، ٠ ) .

$$z_{\alpha} = \sigma_1 = \sigma_2 = 40 \quad \text{و } z_{\beta} = \sigma_1 = \sigma_2 = 20$$

$$\text{اذن } z = \frac{40 \times 2}{20} = \frac{\sigma^2}{n} \quad \text{ومنا هو الخطأ المعيارى للإحصاءة}$$

الفرض الصفري  $F_0$  هو  $\mu_1 = \mu_2$   
 الفرض الآخر  $F_1$  هو  $\mu_1 \neq \mu_2$  (اختبار ذو جانبيين) ،  $\alpha = 0.05$   
 نظراً لأن الاختبار ذو جانبيين فإن منطقة رفض الفرض الصفري تتألف من منطقتين  
 عند ذيلي التوزيع يحددهما العددين  $t_{\alpha/2}$  ،  $t_{1-\alpha/2}$  بحيث  
 $t_{\alpha/2} = (t_1 - \bar{t}_1) / \sqrt{s^2 / n_1} > t_{\alpha/2}$  و  $t_{1-\alpha/2} = (t_2 - \bar{t}_2) / \sqrt{s^2 / n_2} < -t_{\alpha/2}$   
 على أن يحسب كلا الاحتمالين على أساس التسليم بصحة الفرض الصفري  
 $\mu_1 = \mu_2$  أى من الإحصاءة

$$T = \frac{(\bar{t}_1 - \bar{t}_2) - (\mu_1 - \mu_2)}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(\bar{t}_1 - \bar{t}_2) - \text{صفر}}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

التي تتبع التوزيع المعتدل المعياري : مع (1 ، 0) .

$$\text{لدينا } t = \frac{(\bar{t}_1 - \bar{t}_2) - (\mu_1 - \mu_2)}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(\bar{t}_1 - \bar{t}_2)}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} < -t_{\alpha/2} \text{ و } \frac{(\bar{t}_1 - \bar{t}_2)}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} > t_{\alpha/2} \text{ لـ } (t_{\alpha/2} < -t_{\alpha/2})$$

$$= 0.025$$

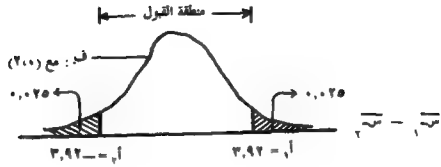
من جدول المساحات أسفل المنحنى المعتدل نجد أن

$$\frac{1}{n_1} = 1.96 \text{ ومنها } \frac{1}{n_2} = 3.92$$

من التماثل نجد أن :

$$\frac{1}{n_1} = 1.96 \text{ ومنها } \frac{1}{n_2} = 3.92$$





الشكل (٧-٧)

وإذن تتألف المنطقة التي نرفض فيها الفرض الصفري ف :  $\mu_1 = \mu_2$  عند المستوى  $0.05$  من اتحاد المنطقتين  $(\bar{x}_1 - \bar{x}_2) < -3.92$  و  $(\bar{x}_1 - \bar{x}_2) > 3.92$  أى من المنطقة التي يأخذ فيها الفرق بين الوسطين الحسابيين للعينات من الحجم  $20$  قيمة تزيد عن  $3.92$  أو قيمة تقل عن  $-3.92$  واحتمال هاتين المنطقتين معا تعبر عنه مساحتي الجزئين المظللين بالشكل (٧-٧) ، وتكون منطقة القبول هي المنطقة المكملة لهاتين المنطقتين أى المنطقة التي تحددها الفترة  $(-3.92, 3.92)$  وتحدد قاعدة الاختبار كالآتي :

« إذا وقع الفرق بين متوسطى عيّنتين مستقلتين من الحجم  $20$  خارج المنطقة  $\bar{x}_1 - \bar{x}_2 : \{ -3.92 \leq \bar{x}_1 - \bar{x}_2 \leq 3.92 \}$  نرفض الفرض الصفري ف. عند مستوى الدلالة  $0.05$  وإلا نقبل ف.  $0.05$  »

(ثانياً)

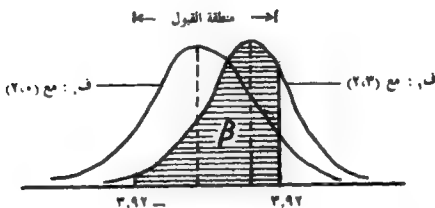
لحساب قيمة  $\beta$  نحسب احتمال وقوع قيم المتغير  $\bar{x}_1 - \bar{x}_2$  في منطقة القبول على أساس أن الفرض الصفري خاطيء والفرض الآخر  $\mu_1 - \mu_2 = 3$  هو الصحيح .

$$\beta = L = P(-3.92 \leq \bar{x}_1 - \bar{x}_2 \leq 3.92) \text{ حيث } \bar{x}_1 - \bar{x}_2 : \text{ مع } (3, 2)$$

$$J = \left( \frac{3 - 3,92}{2} \leq \frac{3 - \bar{x}_1 - \bar{x}_2}{2} \leq \frac{3 - 3,92}{2} \right)$$

من جدول المساحات  $J = (3,46 \leq \epsilon \leq 3,68) = 0,68$

وإذن قوة الاختبار عندما  $\mu_1 - \mu_2 = 3$  هي  $1 - \beta = 0,32$  - انظر الشكل (٨ - ٧).



الشكل (٨ - ٧)

التوزيع المثل للفرض الصفري والتوزيع المثل للفرض الآخر  
(مساحة الجزء المظلل تعبر عن الاحتمال  $\beta$  للمعطى من النوع الثاني)

(ثالثاً)

لإيجاد الحد الأعلى لحجم العينة الذى يضمن أن تكون  $\beta = 0,10$  نستخدم نفس المعادلة (١) السابق تقديمها وهي :

$$\frac{f}{\epsilon_1 - \epsilon_2} = 0,20$$

ويستلزم الأمر هنا أن تكون العيتان مستقلتين ومن نفس الحجم .

$$\frac{8.0}{\sqrt{2}} = \frac{10.2}{\sqrt{2}} = 0.2 \text{ خ. لدينا}$$

$$3 = 0 - 3 = \text{ف.}$$

$$\text{ل (} \epsilon < 1 \text{)} = \alpha = 0.025 \text{ لأن الاختبار ذو جانبيين}$$

$$\text{وإذن } \epsilon = 1.96$$

$$\text{ل (} \epsilon \geq 1 \text{)} = \beta = 0.10 \text{ وإذن } \epsilon = 1.04$$

$$\therefore 1 = \frac{2}{3} = \frac{3}{(1.04) - 1.96} = \frac{8.0}{\sqrt{2}}$$

$$8.0 = \sqrt{2}$$

أى أنه يكفى أخذ عيتين مستقلتين حجم كل منهما 8.0 لكى نضمن أن تكون  $\beta = 0.10$  (تحقق من ذلك بأخذ  $\sqrt{2} = 8.0$  وإثبات أن  $1.96 = 1$  و  $1.96 - 0.1492 = \beta$  وأن  $0.1492 = \beta$ )

### (٧ - ٥) حساب قيمة $\beta$ فى اختبار النسبة :

فى البنود الثلاثة السابقة كنا نتناول الأوساط الحسابية لعينات من مجتمعات معتدلة أو معتدلة تقريبا . على أن المنطق الذى استخدمناه فى حساب الاحتمال  $\beta$  للخطأ من النوع الثانى وحساب قوة الاختبار ينطبق على أى مقاييس أخرى . وفى المثال الآتى نتناول نسبة وقوع حدث ما فى مجتمع ما .

### مثال (٧ - ٦) :

بينت الخبرة أن معدل الشفاء من مرض معين بواسطة علاج قياسي 60٪ . ابتكر علاج جديد يظن أنه أفضل من العلاج القياسي . بين كيف تختبر عند مستوى

الدلالة ٠,٠٥ ، ما إذا كان معدل الشفاء بالعلاج الجديد أعلى منه بالعلاج القياسي ، وذلك باستخدام عينة من ١٥ مريضاً بهذا المرض . حدد قوة الاختبار عندما يفترض أن معدل الشفاء بالعلاج الجديد ٧٠٪ .

الحل :

إن جودة العلاج تقاس بقيمة المتغير  $\pi$  الذى يعبر عن عدد المرضى الذين شفوا في عينة من الحجم ١٥ . وإذا اعتبرنا أن العينة عشوائية ذات وحدات مستقلة فإن المتغير  $\pi$  يكون له توزيع ذى الحدين دليلاً  $\pi$  ، حيث  $\pi = ١٥$  ،  $\pi$  بارامتر مجهول يعبر عن احتمال الشفاء لأى مريض . راجع البند (٣ - ٣) .

ولبحث أفضلية العلاج الجديد ، علينا أن نقارن بين الفرضين الآتيين :

الفرض الصفري  $H_0$  :  $\pi = ٠,٦$  ( لا يوجد فرق في معدل الشفاء بين نوعي العلاج )

الفرض الآخر  $H_1$  :  $\pi < ٠,٦$  ( اختبار ذو جانب واحد )

إذا كان الفرض الصفري صحيحاً ، يكون للمتغير  $\pi$  توزيع ذو حدين دليلاً  $\pi = ١٥$  ،  $\pi = ٠,٦$  ويمكننا حينئذ إيجاد توزيع احتمال هذا المتغير بالحساب المعتاد ( أى من دالة الكتلة ) أو باستخدام الجدول (٣) في ذيل هذا الكتاب مع أخذ  $\pi = ١٥$  ،  $\pi = ٠,٦$  فنجد التوزيع الذى ننقله في الجدول (٧ - ٢) الآتى .

نظراً لأن الاختبار ذو جانب واحد هو الجانب الأيمن فإن منطقة رفض الفرض الصفري هي المنطقة التى يأخذ فيها المتغير  $\pi$  قيمة تزيد عن العدد  $\alpha$  (  $\pi < \alpha$  ) لا تزيد عن  $\alpha = ٠,٠٥$  ، ولإيجاد القيمة الحرجة  $\alpha$  التى نحدد التوزيع من العين نخرج بضعة قيم مستعینين بالجدول (٧ - ٢) كالآتى :

$$\pi = ( \pi < ١٤ )$$

$$\pi = ٠,٠٠٥ = ( \pi < ١٣ )$$

$$ل (س < ١٢) = ٠,٠٢٢ + ٠,٠٠٥ = ٠,٠٢٧$$

$$ل (س < ١١) = ٠,٠٦٣ + ٠,٠٢٧ = ٠,٠٩٠$$

وهذا الاحتمال يزيد عن ٠,٠٥

إذن القيمة الحرجة  $\alpha = ١٢$  وتحدد قاعدة الاختبار كالآتي :

« إذا كان عدد المرضى الذين شفوا في عينة من ١٥ مريضاً يزيد عن ١٢ نرفض الفرض الصفري ف. أن  $ح = ٠,٦$  عند مستوى الدلالة ٠,٠٥ وإلا نقبل ف. »

الجدول (٧-٢)

توزيع الاحتمال لمطير ذي حدين : حد (١٥ ، ٠,٦)

س	ل	س	ل
عدد حالات الشفاء	احتمال هذا العدد	عدد حالات الشفاء	احتمال هذا العدد
٠	_____	٨	٠,١٧٧
١	_____	٩	٠,٢٠٧
٢	_____	١٠	٠,١٨٦
٣	٠,٠٠٢	١١	٠,١٢٧
٤	٠,٠٠٧	١٢	٠,٠٦٣
٥	٠,٠٢٤	١٣	٠,٠٢٢
٦	٠,٠٦١	١٤	٠,٠٠٥
٧	٠,١١٨	١٥	_____
			٠,٩٩٩

لحساب قوة الاختبار عندما يفترض أن  $\pi = 0.7$ ، نحسب احتمال وقوع قيم المتغير  $x$  في منطقة القبول وهي  $x \geq 12$  تحت هذا الفرض أى على أساس أن للمتغير  $x$  توزيعاً ذا حدين دليلاً  $10, 0.7$ .

احتمال الخطأ من النوع الثانى  $\beta = P(x < 12 | \pi = 0.7)$  حيث  $x$  : حد  $(10, 0.7)$   
 $1 - P(x < 12 | \pi = 0.7)$

من الجدول (3) بذيل هذا الكتاب وبأخذ  $n = 10$  و  $\pi = 0.7$  نجد أن :

$$\beta = 1 - (0.92 + 0.31 + 0.005) = 0.765$$

$$0.872 = 0.87 \text{ تقريباً}$$

$$\text{وإذن قوة الاختبار} = 1 - \beta = 0.87 = 0.87$$

وبلاحظ أن قوة الاختبار ضعيفة مما يدعونا إلى الشك في قدرة التجربة على التمييز بين معدل الشفاء في العلاجين القياسى والجديد . وينبغي حيثئذ العمل على زيادة هذه القدرة وذلك بزيادة حجم العينة .

حل آخر :

فى هذا المثال يمكننا استخدام تقريب التوزيع المعتدل لتوزيع ذى

الحدين - راجع البند (4-6) مع ملاحظة أن  $n\pi = 10 \times 0.6 = 6 < 9$

كما أن  $n(1-\pi) = 10 \times 0.4 = 4 < 9$  ، ويقترب توزيع ذى الحدين

حد  $(10, 0.6)$  من توزيع معتدل متوسطه  $\mu = n\pi = 6$  وانحرافه المعياري  $\sigma$

$$= \sqrt{n\pi(1-\pi)} = \sqrt{10 \times 0.6 \times 0.4} = 1.549$$

$$z = \frac{12 - (6 - 0.5)}{1.549}$$

بالتقريب توزيع معتدل معيارى : مع  $(0, 1)$  .

لإيجاد القيمة الحرجة لدينا :

$$J = (S < 1) = (E) = \frac{1 - (0,5 - 1)}{1,897} = 0,05 \text{ فرضا}$$

من جدول المساحات أسفل المنحنى المعتدل المعيارى نجد أن

$$1,64 = \frac{0,5 - 1}{1,897} \quad \text{ومنها } 12,611 =$$

واذن نرفض الفرض الصفري أن  $H = 0,6$  عند مستوى الدلالة  $0,05$  إذا كان عدد المرضى الذين شفوا في عينة من 15 مريضاً يزيد عن 12. وهذه هي النتيجة التى توصلنا إليها بالحل الأول. كذلك :

$$\beta = J = (S \geq 12) \quad \text{حيث } S : \text{حد } (0,7, 15)$$

وبالتقريب بتوزيع معتدل وسطه الحسابى  $u = H = 0,7 \times 15 = 10,5$

$$\text{وانحرافه المعيارى } \sigma = \sqrt{H \cdot (1 - H)} = \sqrt{0,3 \times 0,7 \times 15} = 1,7748$$

نجد أن :

$$\beta = J = (S \geq 12) = (E) = \frac{(10,5 - (0,5 + 12))}{1,7748} \geq$$

$$J = (E \geq 1,13) = 0,8708 = 0,87 \text{ تقريبا}$$

$$0,13 = 0,87 - 1 =$$

## تمارين (٧)

(١) أخذت عينة عشوائية حجمها  $n = 50$  من مجتمع معتدل تباينه  $10,21$ .

بين كيف نختبر عند مستوى الدلالة  $0,05$  الفرض الصفري أن متوسط المجتمع

$\mu = 40,5$  ضد الفرض  $\mu \neq 40,5$ . أوجد قوة الاختبار عندما نفترض أن

$$\mu = 40,5.$$





## الفصل الثامن

### تحليل التباين وتصميم التجارب

#### ANALYSIS OF VARIANCE & DESIGN OF EXPERIMENTS

#### ( ٨ - ١ ) التحليل الإحصائي وتصميم التجارب :

يميل بعض الباحثين التجريبيين إلى تصميم تجاربهم وتنفيذها ، وبعد الانتهاء من الحصول على بيانات ينظرون في تحليل هذه البيانات إحصائياً . وهذا خطأ كبير لأن إغفال الجانب الإحصائي أثناء وضع التصميم غالباً ما يؤدي إلى اختيار تصميم خاطيء لا تستخلص منه أية نتائج يعتد بها . وعلى العكس من ذلك ، إذا أخذ الجانب الإحصائي بعين الاعتبار ، فإنه لا يعاون فقط على تحليل البيانات تحليلًا علمياً سليماً بل يسهم بشكل أساسي في اختيار التصميم الأكثر كفاءة أى الذى يعطى أكبر قدر من المعلومات والنتائج بأدنى حد من الجهد التجريبي ، وهو بالإضافة إلى ذلك يقلل من مصادر أخطاء التجريب ويوضح طريقة تقدير هذه الأخطاء . فخطوة التحليل الإحصائي هي جزء رئيسي من تصميم التجربة . وتقتضي هذه الخطوة مراعاة عدة مبادئ لعل أهمها ما يلي :

#### RANDOMNESS

#### ( أولاً ) العشوائية :

إن التقنية الإحصائية للتجريب تقتضي تطبيق مبدأ العشوائية في كل ما يتعلق بالتجربة منعاً لأى تحيز من أى نوع ، وكوسيلة للتصدى لمجموعة العوامل الثانوية

التي نعجز عن حصرها أو حساب التأثير الطفيف الذي يحدثه كل منها وبالتالي نعجز عن التحكم فيها تجريبياً .

(أ) فالعينة التي تختار من المجتمع ينبغي أن تكون عشوائية ، فتكون مسحوبة بحسب خطة تضمن عدم وجود تحيز من أى نوع قد يؤثر في عملية الاختيار .

(ب) وإذا قسمت هذه العينة إلى أقسام لتطبيق أنواع مختلفة من المعالجات على هذه الأقسام ينبغي أن يكون هذا التقسيم عشوائياً لكي يتوفر لكل وحدة من وحدات العينة نفس الفرصة لتلقي أى من هذه الأنواع .

(ج) كما أن توزيع نوع ما لمعالجة ما على وحدات قسم ما ينبغي أن يكون عشوائياً خاصة من حيث الترتيب الزمني .

وبالنسبة لتقسيم العينة هناك طرق تكفل عشوائية هذا التقسيم ، ومن هذه الطرق ما يلي .

### (١) رمى قطعة معدنية من العملة :

نفرض مثلاً أننا نريد تقسيم ٢٠ حشرة إلى ٤ أقسام يحتوى كل منها على ٥ حشرات لكي تتلقي الحشرات التي تدخل في قسم ما واحداً من ٤ معالجات مختلفة ا ، ب ، ح ، د . نرقم الحشرات من ١ إلى ١٢٠ . نأخذ كل حشرة على حدة ونرمى قطعة منتظمة من العملة مرتين عشوائياً ( أو نلقي قطعتين متميزتين من العملة مرة واحدة ) . نحدد القسم الذي تدخل فيه الحشرة بحسب خطة كالاتية :

الرمية الأولى	الرمية الثانية	القسم ( المعالجة )
صورة	صورة	ا
صورة	كتابة	ب
كتابة	صورة	ح
كتابة	كتابة	د

فإذا أخذنا الحشرة رقم (١) وظهرت كتابة في الرمية الأولى وصورة في الرمية الثانية فإن هذه الحشرة تدخل القسم حـ أى تتلقى المعالجة حـ وهكذا بالنسبة للحشرات جميعاً . وإذا حدث أن امتلأ أحد الأقسام ( بخمس حشرات ) وجاءت رمية لهذا القسم نلقى هذه الرمية ونعيد الرمي . إن هذه الخطة تكفل أن تكون النواتج الممكنة من رمي العملة مرتين وهى ( صورة ، صورة ) و ( صورة ، كتابة ) و ( كتابة ، صورة ) و ( كتابة ، كتابة ) متساوية الاحتمال إذ من الواضح أن احتمال كل منها يساوى  $\frac{1}{4}$  بشرط أن تكون العملة منتظمة والرمي عشوائياً . وبهذا يكون لكل حشرة نفس الفرصة لتلقى أى من المعالجات الأربع . وهذا يعني توفر شرط عشوائية التقسيم . نلاحظ أنه بالنسبة للحشرة الأخيرة لا نكون بحاجة إلى رمي قطعة العملة .

## (٢) رمي حجرة نرد :

في المثال السابق يمكن أن نستخدم خطة أخرى كالآتية :

نرمي حجرة نرد منتظمة عشوائياً . إذا ظهرت نقطة واحدة ندخل الحشرة في القسم ١ وإذا ظهرت نقطتان ندخلها في القسم ب وإذا ظهرت ٣ نقط ندخلها في القسم حـ وإذا ظهرت ٤ نقط ندخلها في القسم د . أما إذا ظهرت ٥ أو ٦ نقط فلا تحسب ويعاد الرمي . نلاحظ هنا أيضاً أن النواتج الستة متساوية الاحتمال فاحتمال كل منها يساوى  $\frac{1}{6}$  .

## (٣) استخدام ورق اللعب :

حين يكون عدد الأقسام المطلوبة كبيراً يحسن استخدام ورق اللعب . نفرض أننا نريد تقسيم ٦٠ حشرة إلى ١٢ قسماً بحيثوى كل منها على ٥ حشرات لكى تتلقى الحشرات التى تدخل فى قسم ما واحداً من ١٢ معالجة مختلفة . نستخدم مجموعة من ورق اللعب بعد استبعاد الملوك الأربعة فيكون لدينا ٤٨ ورقة . نحدد خطة كالآتية :

الواحد للقسم الأول والاثنين للقسم الثاني ، ... ، ... والعشرة للقسم العاشر والولد للقسم الحادى عشر والبنث للقسم الثاني عشر . نأخذ كل حشرة على حدة ونخلط الورق جيداً ثم نقطعه عشوائياً فيكون العدد المقطوع هو الذى يحدد القسم الذى تدخل فيه الحشرة . نلاحظ أن احتمال ظهور أى من الحالات الاثني عشر يساوى  $\frac{1}{12} = \frac{1}{48}$  .

إن الطرق سابقة الذكر هى مجرد أمثلة على طرق التقسيم العشوائى ويمكن للباحث على ضوء هذه الأمثلة أن يتدع طرقاً أخرى كثيرة ، وذلك إضافة إلى إمكانية استخدام جداول الأعداد العشوائية المشار إليها بالبند ( ١ - ٢ ) .

## ( ثانيا ) الاستقلال : INDEPENDENCE

يقتضى التحليل الإحصائى ، خاصة في تحليل التباين ، افتراض استقلال أخطاء التجريب عن بعضها واستقلال المعالجات عن أخطاء التجريب ، فإذا كانت الأخطاء الكبيرة مثلاً مرتبطة بمعالجة معينة فإن استخدام تقدير شامل لخطأ التجريب ، وهو الإجراء المتبع عادة ، لا يكون إجراء سليماً يعتمد عليه في اختبارات الدلالة . على أن تطبيق مبدأ العشوائية سابق الذكر يضمن إلى حد كبير تحقيق هذا الافتراض . كما يسهم في تحقيقه استخدام مجموعات من المشاهدات ( مأخوذة من سلسلة من التجارب ) بدلا من استخدام مجموعة واحدة من المشاهدات .

## ( ثالثا ) النموذج الإحصائى : STATISTICAL MODEL

كما يقتضى التحليل الإحصائى وضع نموذج يرشدنا إلى الأسس الإحصائية التى تحدد أسلوب هذا التحليل ، وتنعكس حدوده تأثيرات العوامل أو المتغيرات التى تدخل في التجريب . وترتبط بكل نموذج افتراضات خاصة تتعلق بتوزيعات هذه المتغيرات أو باستقلالها أو بالصورة الرياضية التى يأخذها النموذج لوصف وحدات التجريب ... وكلما كان النموذج ناجحاً في تصوير التجربة الفعلية كلما كانت النتائج التى نحصل عليها من تحليل البيانات أكثر صدقاً .

## ( رابعاً ) مسائل أخرى :

- ينبغي أن يجيب تصميم التجربة على تساؤلات عدة منها :
- ( أ ) ما هو الحجم المناسب للعينة ؟
- ( ب ) متى يتعين تكرير التجربة برمتها ؟
- ( ج ) متى نحتاج إلى إدخال مجموعة مراقبة ؟ control group
- ( د ) ما الطريقة العملية لتطبيق مبدأ العشوائية ؟
- ( هـ ) ما مدى الدقة والضبط اللازمين في عملية القياس ؟

## ( ٨ - ٧ ) تحليل التباين :

كثيراً ما نلاحظ وجود اختلاف في قيم متغير ما لا نعرف سببه أو مصدره ولا نستطيع التحكم فيه . ومن أمثلة ذلك الاختلاف المشاهد في الزيادة الشهرية في أوزان مجموعة من الماشية حتى لو وضعت في ظروف واحدة وتحت نظام غذائي مشترك ، كذلك الاختلاف المشاهد في نمو وحدات نبات مزروع في حقل تحت نفس الظروف ... إن مثل هذا الاختلاف نصفه بأنه اختلاف عشوائي .

على أننا في كثير من التجارب ندخل سبباً إضافياً للاختلاف في قيم المتغير نعلم مصدره ، فمثلاً قد نقسم مجموعة الماشية إلى عدة أقسام يتلقى كل منها نظاماً مختلفاً للتغذية ، أو قد يقسم الحقل إلى عدة أحواض يتلقى كل منها نوعاً مختلفاً من المخصبات أو طرقاً مختلفة للرعى ونقول حينئذ أننا أدخلنا عاملاً factor معيناً في التجربة . والعامل هو متغير نوعي يتألف من عدد من المعالجات treatments أو التقسيمات المرتبطة تسمى مستويات العامل levels فالمعامل « نظام التغذية » قد يتكون من ٤ مستويات والعامل « نوع المخصب » قد يتكون من ٣ مستويات وهكذا ..

ومن الواضح أن الهدف من إدخال العامل معرفة ما إذا كانت المستويات المختلفة ( لنظام التغذية مثلاً ) تحدث تأثيرات مختلفة في قيم المتغير ( الزيادة في الوزن )

وهذا هو الغرض الذى تستخدم من أجله عملية تحليل التباين . وتؤسّس هذه العملية على تصميم تجربة تمكّننا من أن نفصل الاختلاف الذى سببه العامل عن الاختلاف العشوائى ، فإذا ظهر لنا أن ذلك الاختلاف جوهرى حكمنا بأن عامل التقسيم هو عامل مؤثر فى قيم المتغير وتصلدنا بعد ذلك للمقارنة بين مستويات هذا العامل .

فتحليل التباين هو عملية نستطيع بواسطتها أن نحلل الاختلاف الكلى المشاهد فى مجموعة من البيانات إلى مركبتين أو أكثر يرجع كل منها إلى عامل أو مصدر مستقل ، وإذا كانت هذه البيانات من عينة عشوائية مأخوذة من مجتمع معتدل ذى تباين معين فإن كلا من هذه المركبات يعطى تقديراً مستقلاً لهذا التباين ، والتقدير الناتجة يمكن اختبار تجانسها بواسطة اختبار ف .

وفى التجارب التى نبحت فيها تأثير عامل واحد - كما فى الأمثلة سابقة الذكر - يحلل الاختلاف الكلى إلى مركبتين مستقلتين إحداها تناظر هذا العامل والأخرى تناظر الاختلاف العشوائى . وفى التجارب التى نبحت فيها تأثير عاملين مستقلين ، مثلاً نوع الغذاء كأحد العاملين وكمية الغذاء كعامل ثان ، نحلل الاختلاف الكلى إلى ثلاث مركبات مستقلة اثنتان منهما تناظران العاملين والثالثة تناظر الاختلاف العشوائى ، وهكذا فى حالة وجود أكثر من عاملين .

وفى بحثنا عن كيفية تحليل التباين نحتاج إلى المصطلحات والتعاريف المبينة فى البند التالى .

## ( ٨ - ٣ ) مصطلحات وتعريف :

( ١ ) مجموع المربعات ( ٢ ) :  $\text{SUM OF SQUARES (SS)}$

إذا كان لدينا مجموعة من القيم  $s_1, s_2, \dots, s_n$  فإن مجموع مربعات

انحرافات هذه القيم عن وسطها الحسابي  $\bar{x}$  يسمى اختصاراً بمجموع المربعات ونرمز له بالرمز  $\sum (x_i - \bar{x})^2$  ، أي أن :

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$$

$$\sum x_i^2 - n\bar{x}^2 = \sum (x_i - \bar{x})^2 \quad \text{حيث } \sum (x_i - \bar{x})^2 = \text{مجموع القيم}.$$

ويتخذ  $\sum (x_i - \bar{x})^2$  كمقياس للاختلاف variation في هذه القيم . ويلاحظ أن قيمة  $\sum (x_i - \bar{x})^2$  لا تتغير إذا جمعنا أو طرحنا عددا ثابتا من جميع القيم .

(ب) درجات الحرية (  $\nu$  ) أو ( د. ح )

### DEGREES OF FREEDOM

يستخدم مصطلح « درجات الحرية » في الإحصاء التطبيقي للتعبير عن عدد المقادير المستقلة خطياً في مجموع المربعات . فمثلاً القيم  $x_1, x_2, \dots, x_n$  هي  $n$  من المقادير المستقلة ولكن نظراً لأن  $\sum (x_i - \bar{x})^2 = 0$  فإن المقادير  $(x_1 - \bar{x}), (x_2 - \bar{x}), \dots, (x_n - \bar{x})$  لا تكون مستقلة خطياً لأنه يمكن اشتقاق واحد منها من الآخرين ولذلك يكون لمجموع مربعاتها  $\sum (x_i - \bar{x})^2$  درجات حرية عددها  $n - 1$  .

كقاعدة عامة ، يحسب عدد درجات الحرية لإحصاء ما كالآتي :

= عدد المشاهدات المستقلة المسببة للاختلاف - عدد البارامترات المستقلة التي قدرت من العينة عند حساب هذا الاختلاف .

وفي تحليل التباين نستخدم التعريف الإجرائي الآتي للدرجات الحرية لأي مصدر من مصادر الاختلاف .

= عدد الانحرافات المربعة - عدد النقط ( المحاور ) المستقلة التي أخذت حولها هذه الانحرافات . ( يلاحظ أن عدد النقط أو المحاور هذه هي عدد القيود الخطية التي فرضت على تقدير الاختلاف ) .

### (ج) متوسط المربعات (ع') أو (ط ٢) MEAN SQUARE (MS)

هو خارخ قسة مجموع المربعات علي عدد درجات الحرية أى  $م = م / ل$  ويسمى هذا بالتباين ، غير أن التعبير متوسط المربعات هو تعبير أكثر عمومية .

### (٨ - ٤) التجارب ذوات العامل الواحد :

#### SINGLE FACTOR EXPERIMENTS

في هذه التجارب يكون اهتمامنا منصبا على دراسة عامل واحد فقط ، وليكن نظام التغذية ، من حيث تأثيره على متغير ما وليكن الزيادة في وزن نوع من البقر في مدة م . ولا يغرب عن بالنا هنا إمكانية وجود مصادر أو عوامل أخرى ذات تأثير على هذا المتغير مثل عمر البقر وجنسه ووزنه الأصل .. ولذلك ينبغي أن نعمل على تجميع تأثير هذه العوامل ومنع تداخل هذا التأثير مع التأثير الذى يحدثه العامل الذى ندرسه .

ولتحقيق هذا الغرض يلجأ بعض الباحثين إلى تصميم تجربة يتحكم فيها تحكما كاملا في هذه العوامل فيقوم بتثبيتها عند مستويات عديدة فيختار مجموعة من البقر في نفس العمر ومن نفس الجنس ونفس الوزن .. ويسمح فقط بتغيير عامل التغذية وذلك بتقسيم مجموعة البقر إلى عدة أقسام ومعالجة كل قسم بواحد من مستويات نظام التغذية ، وبهذا يخلو مسئولية أى من تلك العوامل مما قد يظهر من فروق جوهرية بين هذه المستويات . غير أن النتائج التى تسفر عنها هذه الطريقة تكون مشروطة بتوفر الظروف الخاصة التى هيئت لها التجربة من حيث العمر والجنس والوزن .. وقد لا تكون هذه النتائج صحيحة إذا ما تغير أى من هذه الظروف ، وبالتالي لا تعطى التجربة قدرا كافيا من المعلومات التى ينشدها الباحث . ومن ناحية أخرى من الصعب عمليا بل ومن المستحيل أحيانا التحكم في التجربة وضبط تلك العوامل الخارجية عند مستوى مشترك بالدقة الكافية .



ولذلك يفضل الباحثون تصميم تجربة على النقيض من ذلك ، فبدلاً من أن نتحكم في العوامل الخارجية بوضعها عند مستويات خاصة ، نقوم بتعشية هذه العوامل بحيث يكون لكل مستوى من مستويات العامل الذى ندرسه نفس الفرصة للتعرض لها ، وبهذا يحق لنا ضم تأثير هذه العوامل تحت كلمة عامة هى **الاختلاف العشوائى** أو خطأ التجريب أو الصدفة . فإذا كان لدينا ٢٨ بقرة و ٤ مستويات من نظام التغذية نرقم البقر من ١ إلى ٢٨ بصرف النظر عن العمر والجنس والوزن لتقسيم البقر عشوائياً إلى ٤ أقسام يحتوى كل منها على ٧ بقرات لتتلقى واحداً من مستويات نظام التغذية . إن مثل هذا التصميم يسمى **بال تصميم كامل التعشية** completely randomized design وهو ذلك التصميم الذى يبنى لا على أساس التحكم في المصادر الخارجية وإنما على أساس تعشية هذه المصادر بحيث يمكن ضم الاختلافات الناشئة عنها تحت اسم **الاختلاف العشوائى** .

على أن طريقة التعشية في الحماية من تأثير العوامل الخارجية هى عملية مبنية على أساس احتمالى ، فقد تسفر هذه الطريقة عن أن يشتمل أحد الأقسام على ٧ بقرات كلها من الإناث أو كلها من صغار السن - وإن كان هذا أمراً بعيد الاحتمال - وهذا أحد الأسباب التى تجعل بعض الباحثين يميل إلى استخدام تصميم وسط بين التقيضين المذكورين وذلك بالتحكم في بعض العوامل وتعشية البعض الآخر . وسنعود إلى هذا الموضوع في البند (٨ - ٧) تحت عنوان المقارنات التزاوجية.

اعتبر عينة عشوائية حجمها  $n$  مأخوذة من متغير معتدل  $\mu$  وسطه الحسابى  $\mu$  وتباينه  $\sigma^2$  . افترض أن هذه العينة قسمت عشوائياً إلى  $k$  من الأقسام تلقت كل منها واحداً من مستويات عامل ما ( نظام الغذاء - نوع المخصب - طريقة الرى - ... ) وجاءت البيانات كما يلي ، حيث  $\bar{y}_i$  ترمز إلى قيم المتغير الذى ندرسه ( مقدار المحصول مثلاً ) ، و  $\bar{y}$  ترمزان على الترتيب إلى رقم الصف ورقم العمود الذى تقع فيه القيمة  $y_{ij}$  .

الأقسام (المعالجات)							
(١)	(٢)	(٣)	...	(ق)	...	(ك)	
١١ س	٢١ س	٣١ س	...	...	...	١١ س	
١٢ س	٢٢ س	٣٢ س	...	...	...	١٢ س	
١٣ س	٢٣ س	٣٣ س	...	...	...	١٣ س	
...	...	...	...	...	...	...	
...	...	...	...	...	...	...	
١ ر	٢ ر	٣ ر	...	...	...	١ ر	
...	...	...	...	...	...	...	
١٥ س	٢٥ س	٣٥ س	...	...	...	١٥ س	
ن ق	ن ٢	ن ٣	...	ن ق	...	ن ٤	ن = محن ق
ق ٢	ق ٢	ق ٢	...	ق ٢	...	ق ٢	م = محم ق
س ق	س ق	س ق	...	س ق	...	س ق	س = محم/ن

نر ترمز إلى عدد قيم المتغير في القسم  $q$  ( $q = 1, 2, \dots, k$ ) ،  
 ن ترمز إلى العدد الكلى لقيم المتغير ،  
 م ترمز إلى مجموع قيم المتغير في القسم  $q$  ،  
 س ترمز إلى متوسط قيم المتغير في القسم  $q$  ، س ترمز إلى المتوسط العام .  
 المطلوب بحث ما إذا كان المجتمع متجانساً بالنسبة لهذا التقسيم أى ما إذا كانت  
 مستويات هذا العامل تحدث تأثيرات متساوية في قيم المتغير س .

## ( ٨ - ٤ - ١ ) النموذج الإحصائي ( النموذج I ) :

كما سبق القول يعتمد التحليل الإحصائي على اختيار نموذج يعبر عن تركيب أى عنصر مشاهد في التجربة ويبرر ما يجري من عمليات مصحوباً بافتراضات يقتضيها البناء الرياضي الذي تقوم عليه عملية التحليل . وسنفترض هنا ما يلي :

(١) المجتمع العام الذي أخذت منه وحدات التجربة هو مجتمع معادل وسطه الحسابي  $\mu$  وتباينه  $\sigma^2$  .

(٢) مجموعات الوحدات في الأقسام ١ ، ٢ ، ... ،  $k$  التي تلقت مستويات مختلفة من عامل التجربة تشكل عينات عشوائية مستقلة مأخوذة من  $k$  من المجتمعات المعادلة أوساطها الحسابية  $\mu_1$  ،  $\mu_2$  ، ... ،  $\mu_k$  ولها تباين مشترك  $\sigma^2$  .  
 $(\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2)$  .

(٣) أى وحدة مشاهدة  $y_{ij}$  تخضع للنموذج الخطي الآتي :

$$y_{ij} = \mu + x_{ij} \quad r = 1, 2, \dots, n \quad (1)$$

حيث  $\mu$  هو الوسط الحسابي لمجتمع القسم  $q$  ( $q = 1, 2, \dots, k$ ) .  
 $x_{ij} = y_{ij} - \mu$  هو خطأ التجريب بالنسبة للوحدة  $y_{ij}$  التي في الصف  $r$  والعمود ( القسم )  $q$  أى أن قيم  $x_{ij}$  تعبر عن الفروق العشوائية بين الوحدات داخل القسم  $q$  ، وسنفترض أنه في مجتمع القيم  $x_{ij}$  تكون هذه القيم مستقلة ويكون هذا المجتمع معتدلاً وسطه الحسابي صفر وتباينه  $\sigma^2$  .

ومن المعتاد أن يكتب النموذج (١) كالآتي :

$$y_{ij} = \mu + \alpha + x_{ij} \quad (2)$$

حيث  $\alpha = \mu - \mu$  = انحراف متوسط مجتمع القسم  $q$  عن متوسط المجتمع العام ، وهذا الرمز يصلح للتعبير عن متوسط أثر المستوى  $q$  لعامل التقسيم .

وسنعتبر أن هذا الأثر ثابت لكل وحدة بالقسم ق وأنه يختلف من قسم إلى آخر ، بمعنى أن كل عنصر من عناصر القسم الأول يتأثر ( بالزيادة أو النقصان ) بمقدار ثابت  $\alpha$  ، وكل عنصر من عناصر القسم الثاني يتأثر بمقدار ثابت  $\alpha$  ، وهكذا ... ولذلك يسمى هذا النموذج بالنموذج ثابت التأثيرات fixed effects model تمييزاً له عن النموذج عشوائى التأثيرات random effects model حيث لا تتأثر عناصر الأقسام بمقادير ثابتة بل بمقادير عشوائية . وسوف نتناول هذا النموذج فى البند ( ٨ - ١٤ ) .

إن هذا النموذج هو الأساس الذى يبنى عليه التحليل ، فهو أولاً يفترض أن أى قيمة مشاهدة  $y_{ij}$  يمكن تجزئتها إلى مركبات تعزى إلى مصادر منطقية متميزة تبيينها كالآتى :

من (٧) :  $y_{ij} = \mu + (\mu_j - \mu) + (\bar{y}_{i.} - \mu_j)$  (٣)  
ومن هذه الصيغة نرى أن  $\bar{y}_{i.}$  وهى القيمة المشاهدة فى الصف  $i$  من القسم  $j$  تساوى المتوسط العام للمجتمع + تأثير يرجع إلى المعالجة التى تلقتها وحدات القسم  $j$  + تأثير عشوائى داخل القسم  $j$  .

كما أن هذا النموذج يحدد العلاقة بين الاختلافات الناشئة عن مختلف المصادر أو العوامل المؤثرة فى عملية التجريب :

من (٣) :  $y_{ij} - \mu = (\mu_j - \mu) + (\bar{y}_{i.} - \mu_j)$   
أو (٤)  $(y_{ij} - \bar{y}_{i.}) = (\mu_j - \mu) + (\bar{y}_{i.} - \mu_j)$   
وذلك بوضع المتوسطات  $\bar{y}_{i.}$  ، الناتجة فى العينة بدلا من المتوسطات النظرية المجهولة  $\mu_j$  ،  $\mu$  . بتربيع كل من الطرفين والجمع على جميع قيم  $i, j$  ، نتنتج المتطابقة الآتية :

$$\sum_j \sum_i (y_{ij} - \bar{y}_{i.})^2 = \sum_j n_j (\mu_j - \mu)^2 + \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 \quad (٥)$$



من هذا نحصل على متوسط المربعات لكل من المركبتين كالآتي :

$$\sigma^2 = \frac{22}{2} \text{ ( بين الأقسام ) } / (ك - 1)$$

$$\sigma^2 = \frac{22}{2} \text{ ( داخل الأقسام ) } / (ن - ك)$$

تحت الفروض سابقة الذكر نستطيع أن نثبت رياضياً أن  $\sigma^2$  ،  $\sigma^2$  هما تقديران مستقلان لتباين المجتمع  $\sigma^2$  . غير أن التقدير الثاني ( $\sigma^2$ ) هو تقدير غير متحيز لتباين المجتمع بمعنى أن متوسط مثل هذه التقديرات يكون على المدى البعيد مساوياً للتباين  $\sigma^2$  ، أما التقدير الأول ( $\sigma^2$ ) فهو تقدير متحيز لتباين  $\sigma^2$  إذ أن متوسط مثل هذه التقديرات على المدى البعيد  $\sigma^2 + \frac{1}{1 - \frac{1}{n}} \mu$  أي يزيد عنه ولا يكون مساوياً له إلا إذا كانت المتوسطات  $\mu$  ،  $\mu$  ،  $\mu$  ،  $\mu$  ،  $\mu$  متساوية جميعاً . وعلى ذلك فإن أى فرق جوهري بين التقديرين  $\sigma^2$  ،  $\sigma^2$  لا ينتج إلا من وجود فرق جوهري بين هذه المتوسطات .

#### (٨ - ٤ - ٢) اختبار تجانس المجتمع بالنسبة لعامل التقسيم :

إن الفرض الصفري هو أن المجتمع متجانس بالنسبة لعامل التقسيم وهذا يتضمن أن يكون للمجتمعات المحتدة للأقسام نفس الخواص الإحصائية ، وبصفة خاصة :

$$\sigma^2 = \sigma^2 = \dots = \sigma^2 = \sigma^2 , \mu = \dots = \mu = \mu$$

وإذا كان هذا الفرض صحيحاً ، ونظراً لأن  $\sigma^2$  ،  $\sigma^2$  هما تقديران مستقلان للتباين  $\sigma^2$  لمجتمع محتدل فإن نسبة التباين

$$F = \frac{\sigma^2}{\sigma^2} \quad (٦)$$

يكون توزيعها مطابقاً لتوزيع المتغير ف يدرجتي حرية (ك - ١ ، ن - ك) - راجع البند (٦ - ٨) . وعلى ذلك يمكن استخدام اختبار ف للحكم على هذا التجانس وبالتالي للحكم على تساوى تلك المتوسطات .

أما الفرض الآخر ف، فهو أن الاختلاف الناشئ عن عامل التقسيم ( بين المتوسطات ) أكبر مما نتوقعه من اختلاف عشوائي في عينة من مجتمع متجانس بالنسبة لعامل التقسيم ، ولذلك فإن هذا الاختبار يكون دائماً ذا جانب واحد .

### ( ٨ - ٤ - ٣ ) طريقة مختصرة لحساب الاختلاف :

لتسهيل حساب القيم العددية لمجاميع المربعات الثلاثة المبينة بالمطابقة (٥) نستخدم الصيغ الآتية التي يمكن برهنتها رياضياً .

$$(١) \text{ (الكل) } \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{حيث } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} = ١ - ٥$$

$$(٢) \text{ (بين الأقسام) } \sum_{j=1}^k \sum_{i=1}^n x_{ij}^2 - \frac{(\sum_{j=1}^k \sum_{i=1}^n x_{ij})^2}{n} = \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - \bar{x}_{.j})^2 \quad \text{حيث } \bar{x}_{.j} = \frac{\sum_{i=1}^n x_{ij}}{n} = ١ - ٤$$

$$(٣) \text{ (داخل الأقسام) } \sum_{j=1}^k \sum_{i=1}^n x_{ij}^2 - \sum_{j=1}^k \frac{(\sum_{i=1}^n x_{ij})^2}{n_j} = \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - \bar{x}_{j.})^2 \quad \text{حيث } \bar{x}_{j.} = \frac{\sum_{i=1}^n x_{ij}}{n_j} = ١ - ٥$$

وتوضع هذه القيم عادة في جدول يسمى بجدول التباين يأخذ الصورة الآتية :

الجدول (٨ - ٧)

جدول التباين للضارب ذوات العامل الواحد

مصدر التباين	٢ ٢	د ح	تقدير التباين	ف.م
بين الأقسام	(٢)	١ - ٤	$\sum_{j=1}^k \bar{x}_{.j}^2$	$\sum_{j=1}^k \bar{x}_{.j}^2 / ٤$
داخل الأقسام	(٢) - (١)	١ - ٥	$\sum_{j=1}^k \sum_{i=1}^n x_{ij}^2 - \sum_{j=1}^k \bar{x}_{.j}^2$	$\sum_{j=1}^k \sum_{i=1}^n x_{ij}^2 - \sum_{j=1}^k \bar{x}_{.j}^2$
المجموع	(١)	١ - ٥		

### ملاحظة (١) :

يفضل أن تكون حجوم العينات في الأقسام المختلفة متساوية أى  $n_1 = n_2 = \dots = n_k = 1$  مثلاً لأنه في هذه الحالة لا يكون تحليل التباين حساساً للانحرافات الصغيرة عن فرض تساوى التباينات في مجتمعات الأقسام . هذا بالإضافة إلى تسهيل حساب مجاميع المربعات بين الأقسام حيث تكتب كما يلي :

$$S^2_{\text{بين الأقسام}} = \frac{\sum x_i^2}{1} - \frac{(\sum x_i)^2}{n}$$

### ملاحظة (٢) :

لا تتأثر نتائج تحليل التباين بأي حال إذا جمعنا أو طرحنا عدداً ثابتاً من جميع قيم وحدات التجريب .

### مقال (٨ - ١) :

البيانات التي بالجدول (٨ - ٣) نتجت عن تجربة في فسيولوجيا النبات ، وهي تعطى الطول ( بوحدات شفرية ) لمقاطع من نبات البسلة تركت لتنمو في مزرعة نمجية في وجود هرمون الأوكسين ، وكان الهدف من التجربة اختبار تأثير إضافة أربعة أنواع من السكريات على النمو مقاساً بواسطة الطول .

( على فرض أن مبادئ العشوائية والاعتدالية قد روعيت في إجراء التجربة . )

### الحل :

الفرض الصفري  $F$  هو أن المجموع ( المحتدل ) الذى أخذت منه العينة متجانس بالنسبة لعامل التقسيم ( نوع السكر ) أى أن  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$  . والفرض الآخر  $F_1$  هو : على الأقل اثنان من المتوسطات غير متساويين .



جدول (٨ - ٣)

المعطيات					
(١)	(٢)	(٣)	(٤)	(٥)	
٧+٪ جلوكوز + ٢٪ فركتوز + ١٪ جلوكوز + ٢٪ سكروز + مرابحة					
١+٪ فركتوز					
٥٧	٥٨	٥٨	٦٢	٧٥	
٥٨	٦١	٥٩	٦٦	٦٧	
٦٠	٥٦	٥٨	٦٥	٧٠	
٥٩	٥٨	٦١	٦٣	٧٥	
٦٢	٥٧	٥٧	٦٤	٦٥	
٦٠	٥٦	٥٦	٦٢	٧١	
٦٠	٦١	٥٨	٦٥	٦٧	
٥٧	٦٠	٥٧	٦٥	٦٧	
٥٩	٥٧	٥٧	٦٢	٧٦	
٦١	٥٨	٥٩	٦٧	٦٨	
١٠	١٠	١٠	١٠	١٠ = ٥	١٠
٥٩٣	٥٨٢	٥٨٠	٦٤١	٧٠١	٣٠٩٧ = ٢
٥٩,٣	٥٨,٢	٥٨	٦٤,١	٧٠,١	٦١,٩٤ = ٢

$$\begin{aligned}
 & ٢٢ (الكل) = ١٥٧ + ١٥٨ + \dots + ١٦١ + ١٥٨ + \dots + ١٥٨ \\
 & ١٥٩ + \dots + ١٦٢ + ١٦٦ + \dots + ١٧٥ + ١٦٧ + \dots + ١٦٨ - \frac{(٣٠٩٧)}{٥٠} =
 \end{aligned}$$

$$١٣٢٢,٨٢ = ١٩١٨٢٨,١٨ - ١٩٣١٥١ =$$

$$\text{حيث } ٤٦ = ١ - ٥ = ١$$

$${}^2_{(3.97)} - {}^2_{(7.1)} + \dots + {}^2_{(582)} + {}^2_{(593)} = (\text{بين الأقسام})$$

$$1.77,32 = 191828,18 - 19290.50 =$$

$$\text{حيث } \chi^2 = 1 - \text{ك} =$$

$$245.50 = 1.77,32 - 1322.82 = (\text{داخل الأقسام})$$

$$\text{حيث } \chi^2 = \text{هـ} - \text{ك} = 45$$

نضم هذه النتائج في جدول التباين الآتي :

جدول (٨ - ٤)

مصدر الاختلاف	مجموع المربعات	درجات الحرية	تقدير التباين	ف.ح.
بين الأقسام (المعالجات)	1.77,32	٤	٢٦٩,٣٣	٤٩,٣٣
داخل الأقسام (خطأ التجريب)	٢٤٥,٥٠	٤٥	٥,٤٦	
المجموع	١٣٢٢,٨٢	٤٩		

من جدول ف ، وعند درجتي الحرية ٤ ، ٤٥ نجد أن :

ف.٠٠١ = ٢,٥٨ ، ف.٠٠١ = ٣,٧٧ ، ف.٠٠١ = ٥,٥٧

الاستنتاج :

نظرا لأن ف.ح. = ٤٩,٣٣ أكبر بكثير من أى من هذه القيم فإننا نرفض الفرض الصفري عند مستوى عالى من الدلالة ونحكم بأن الأنواع المختلفة من السكريات ليست متساوية في تأثيرها على نمو مقاطع نبات البسلة .

### ملاحظة (٣) :

في نسبة التباين  $F$  نضع  $E'$  دائماً في البسط و  $E''$  في المقام وإذا حدث أن كانت  $E'$  أصغر من  $E''$  أى كانت  $F > 1$  نقبل الفرض الصفري فوراً دون حاجة إلى إيجاد أى قيمة حرجة من الجدول لأن جميع هذه القيم أكبر من الواحد حين تزيد كل من درجتى الحرية عن الواحد .

### ملاحظة (٤) :

حين يكون عدد الأقسام  $k = 2$  تكون نتيجة تحليل التباين مطابقة للنتيجة التي نحصل عليها باستخدام اختبارات وتكون  $F(1, n-1) = t(n-1)$  .  
ولذلك يمكن أن نعتبر أن اختبارات الفرق بين متوسطى مجتمعين معتلين هو حالة خاصة من اختبار  $F$  .

### (٨ - ٥) المقارنة بين المتوسطات :

في البند السابق أجرينا تحليلاً للتباين للمشاهد في البيانات التي أسفرت عنها التجربة ، غير أن هذا التحليل ليس إلا الخطوة الأولى لدراسة نتائج التجربة وينبغى أن يستكمل بإجراء مقارنات بين بعض أزواج هذه المتوسطات أو بين مجموعات منها . ودراسة هذه المقارنات قد يكون أكثر أهمية من التحليل العام . وهناك نوعان من المقارنات هما :

A priori (or planned) comparisons	(أ) المقارنات القبلية
A postiori (or unplanned) comparisons	(ب) المقارنات اللاحقة

ولعل سبب التمييز بين هذين النوعين هو اختلاف اختبارات الدلالة فهما - كما سيتبين بعد -

## (٨ - ٥ - ١) الاختبارات القبلية :

هى تلك الاختبارات التي كان مخططاً لها أثناء تصميم التجربة ( وقبل إجرائها ) . ففي المثال (٨ - ١) كان مخططاً لاختبار تأثير إضافة السكريات ضد مجموعة المراقبة ، كما كان مخططاً لاختبار ما إذا كانت السكريات النقية ككل ( جلوكوز - فركتوز - سكروز ) تختلف في تأثيرها عن السكريات المختلطة (١٪ جلوكوز + ١٪ فركتوز) .

إن مثل هذه الاختبارات تجرى بصرف النظر عن النتيجة العامة لتحليل التباين أى سواء رفضنا أو قبلنا الفرض الصفري عن تساوى المتوسطات .

والقاعدة التي ننبهها للمقارنة هى نفس القاعدة العامة وليس علينا إلا مراعاة أن نتناول فقط البيانات التي بالأقسام التي نرغب في مقارنتها ، وأن ننسب تقدير التباين الناتج منها إلى تقدير التباين ( داخل الأقسام ) السابق لإنجاده في التحليل العام وهو 'ع' لأن هذا التقدير مبني على جميع ما لدينا من بيانات ( وليس على جزء منها ) فهو أقدر على تقدير تباين المجتمع  $\sigma^2$  .

والأمثلة الآتية هى استكمال لدراسة التجربة التي بالمثال (٨ - ١) .

مثال (٨ - ٢) مقارنة مجموعة المراقبة ضد المجموعات الأخرى :

ابحث ما إذا كانت إضافة السكريات تؤثر في نمو مقاطع البسلة .

الحل :

الفرض الصفري هو أن متوسط مجموعات السكريات الأربعة مجتمعة يساوى متوسط مجموعة المراقبة . نعتبر أن لدينا قسمين يتألف الأول من الأقسام ١ ، ٢ ، ٣ ، ٤ وبها ٤٠ عنصراً ، مجموع قيمها ٢٣٩٦ ، ومتوسطها ٥٩,٩ ، ويتألف

الثاني من قسم المراقبة وبه ١٠ عناصر مجموع قيمها ٧٠١ ومتوسطها ٧٠,١  
ويكون المجموع الكلي لقيم العناصر التي اخترناها لهذه الدراسة ٣٠٩٧ .

$$٢٢ (السكريات ضد المراقبة) = \frac{٢٣٩٦}{٤٠} + \frac{٧٠١}{١٠} - \frac{٣٠٩٧}{٥٠}$$

$$٨٣٢,٣٢ = ٧,٢٢ - ١ = ١$$

$$\text{تقدير التباين بين القسمين} = \frac{٨٣٢,٣٢}{١} = ٨٣٢,٣٢$$

$$١٥٢,٤٤ = \frac{٨٣٢,٣٢}{٥,٤٦}$$

$$\text{من الجدول : ف } ٧,٢٣ = [١٠, ٤٥]$$

نرفض الفرض الصفري عند مستوى الدلالة ٠,٠١ ونظراً لأن متوسط  
السكريات يقل عن متوسط المراقبة نستنتج أن إضافة السكريات يؤخر نمو مقاطع  
نبات البسلة .

مثال (٨ - ٣) مقارنة السكريات النقية ضد السكر الخليط :

قارن تأثير إضافة السكريات النقية ( مجتمعة ) وتأثير إضافة السكر الخليط .

الحل :

الفرض الصفري هو أن متوسط أقسام السكريات النقية معاً يساوي متوسط  
قسم السكريات الخليط .

نعتبر هنا أيضاً أن لدينا قسمين يتألف الأول من الأقسام ١ ، ٢ ، ٤ ككل  
وبها ٣٠ عنصراً مجموع قيمها ١٨١٦ ومتوسطها ٦٠,٥٣ ويتألف الثاني من القسم  
٣ وبه ١٠ عناصر مجموع قيمها ٥٨٠ ومتوسطها ٥٨ ويكون المجموع الكلي لقيم  
العناصر التي اخترناها للدراسة ٢٣٩٦ .

$$٢٢ \text{ (سكريات نقية ضد سكريات خليط)} = \frac{١٨١٦}{٣٠} - \frac{٥٨٠}{١٠} + \frac{٢٣٩٦}{٤٠}$$

$$= ٤٨,١٣ \text{ حيث } ٧ = ٣ - ٢ = ١$$

$$\text{في } ٨,٨٢ = \frac{٤٨,١٣}{٥,٤٦}$$

$$\text{من الجدول : ف. } ١٠,١, ١٠,١, ١٠,١ = ٧,٢٣$$

نرفض الفرض الصفري عند مستوى الدلالة ٠,٠١ ونظراً لأن متوسط مجموعة السكريات النقية أكبر من متوسط مجموعة السكر الخليط نستنتج أن إضافة السكريات النقية يؤخر نمو النبات بدرجة أقل مما يؤخره السكر الخليط .

مثال (٨ - ٤) مقارنة مجموعات السكريات النقية معاً :

ابحث ما إذا كان هناك فرق جوهري بين تأثير السكريات النقية الثلاثة .

الحل :

الفرض الصفري هو عدم وجود فروق بين متوسطات الأقسام الثلاثة . نعتبر هنا أن لدينا ثلاثة أقسام ١ ، ٢ ، ٤ مجموع قيمها ١٨١٦ .

$$٢٢ \text{ (بين الأقسام)} = \frac{١٨١٦}{٣٠} - \frac{٦٤١}{١٠} + \frac{٥٨٢}{١٠} + \frac{٥٩٣}{١٠}$$

$$= ١٩٦,٨٧ \text{ حيث } ٧ = ٣ - ٢ = ١$$

$$\text{تقدير التباين بين الأقسام} = \text{ع} = ٩٨,٤٣٥ = ١٩٦,٨٧ \div ٢$$

$$\text{في } ١٨,٠٦ = \frac{٩٨,٤٣٥}{٥,٤٦}$$

من الجدول نجد أن ف...<sub>(٤٥.٧)</sub> تقع بين ٥,١٨ ، ٤,٩٨ ،  
وإذن نرفض الفرض الصفري عند مستوى الدلالة ٠,٠١ ونستنتج أن السكريات  
النقية تختلف في تأثيرها ، ويبدو أن هذا الاختلاف يرجع إلى السكروز الذى له  
متوسط أعلى بكثير من متوسط النوعين الآخرين .  
نستطيع تلخيص ما توصلنا إليه حتى الآن في الجدول الآتي :

جدول ( ٨ - ٥ )

مصدر الاختلاف	مجموع المربعات	درجات الحرية	تقدير التباين	ف
بين الأقسام	١٠٧٧,٣٢	٤	٢٦٩,٣٣	٢٤٩,٣٣
المزوجة ضد السكريات	٨٣٢,٣٢	$\left\{ \begin{array}{l} ١ \\ ١ \\ ٢ \end{array} \right.$	٨٣٢,٣٢	١٥٢,٤٤
السكريات النقية ضد الخليط	٤٨,١٣		٤٨,١٣	٨,٨٢
بين السكريات النقية داخل الأقسام	١٩٦,٨٧		٩٨,٤٤	١٨,٠٣
	٢٤٥,٥٠	٤٥	٥,٤٦	
المجموع	١٣٢٢,٨٢	٤٩		

### ملاحظة (٥) :

إن تحديد شكل وعدد الاختبارات القبلية يتوقف على التساؤلات التي تطرحها  
المشكلة . على أن هناك تحفظات ينبغي مراعاتها ، فلا يجب أن يزيد مجموع درجات  
الحرية للمقارنات القبلية عن  $k - ١$  حيث  $k$  عدد الأقسام ، ومن الواضح إذن  
أنه من غير المناسب أن نقرر مقدماً إجراء المقارنة بين متوسطات كل زوج من  
الأقسام وهي تتطلب  $\frac{1}{2} = \frac{1}{2}$  كـ (ك - ١) من المقارنات وهذا العدد أكبر  
من  $k - ١$  حين  $k < ٢$  .

وبالإضافة إلى ذلك يفضل أن تختار الاختبارات القبلية بحيث تكون مستقلة ، وذلك لكي تكون المعلومات الناتجة من أى منها ذات قيمة بذاتها وغير متداخلة مع المعلومات الناتجة من أى مقارنة أخرى ، وهذا ما فعلنا في المثال السابق ، إذ أجرينا الاختبارات المستقلة الآتية :

(١) ضد (٢)	ضد (٤)	بدرجات حرية عددها ٢
(١ ، ٢ ، ٤)	ضد (٣)	بدرجات حرية عددها ١
(١ ، ٢ ، ٣ ، ٤)	ضد (٥)	بدرجات حرية عددها ١

وقد أدى هذا الاستقلال إلى أن يكون مجموع مجاميع المربعات في المقارنات الثلاث التي بنيت على ٢ ، ١ ، ١ من درجات الحرية مساوياً لمجموع المربعات بين الأقسام في التحليل العام الذي بني على ٤ درجات حرية . وهذا واضح في الجدول (٨ - ٥) . أى أن مجموع المربعات بين الأقسام قد تمحلل إلى ثلاثة أجزاء منفصلة كل منها هو مجموع مربعات قائم بذاته وله درجات حرية خاصة به .

#### (٨ - ٥ - ٢) الاختبارات البعدية :

هى تلك الاختبارات التي لم يخطط لإجرائها أثناء تصميم التجربة ولكنها تقترح نفسها عند التأمل فيما وصلنا إليه من نتائج بعد إجراء التجربة وتحليل البيانات ، إذ أن هذا التأمل يجعلنا نشبه في وجود فروق جوهرية بين بعض الأقسام مما يستحق البحث والاختبار .

ففي التجربة التي بالمثال (٨ - ١) نشعر بأن هناك فرقاً كبيراً بين متوسط السكروز والمتوسطات الأخرى من السكريات مما يوحى بضرورة اختبار دلالة هذا الفرق . كذلك نشعر بأن الفرق بين متوسط السكروز ومتوسط المراقبة يستحق الاختبار .

إن مثل هذه الاختبارات لا تجرى إلا إذا كانت النتيجة العامة لتحليل التباين



ذات دلالة أى حين يرفض الفرض الصفرى عن تساوى جميع المتوسطات لأنه لو ظهر أن هذه المتوسطات متساوية فإن هذا يتضمن أن الفروق الظاهرة بين أى قسمين لا تكون فروقاً ذات دلالة وبالتالي فإن أى اختبار تجريه لا يضيف جديداً لما علمناه عن دلالة هذه الفروق .

على أن المقارنات البعدية تحتاج لتقرير دلالتها إلى طرق خاصة تختلف عن تلك التي استخدمت في التحليل العام وفي المقارنات القبلية . ذلك لأن الأقسام التي نأخذها للمقارنة ولو أنها مأخوذة من نفس المجتمع العام إلا أننا نقتطعها عمداً من جزء متحيز من التوزيع فهي تفتقد عنصر العشوائية ولا يصبح توزيع الاحتمال الذى أسست عليه عملية اختيار الفروض صالحاً لها . وإذا تناولنا التجربة بالمثل (٨ - ١) وقارناً القسم (٣) الذى أعطى أصغر متوسط والقسم (٥) الذى أعطى أكبر متوسط نكون قد أخذنا الجزء المتطرف الأيسر والجزء المتطرف الأيمن من التوزيع ويكون من السهل ظهور اختلاف كبير بين متوسطيهما حتى ولو كانا من نفس المجتمع . وإذا أردنا الدقة في الحكم فيجب أن نأخذ هذا في الاعتبار وذلك بتصويب تقرير دلالة مثل هذا الفرق .

ويعتمد أحد طرق الاختبارات البعدية على إيجاد قيمة حرجة لمجموع المربعات إذا تعداها مجموع المربعات بين قسمين أو مجموعة من الأقسام يكون الاختلاف بينها ذا دلالة وإلا فهو ليس كذلك . وتوجد هذه القيمة الحرجة كما يلي :

إن الاختبارات البعدية لا تجري كما سبق القول إلا إذا كانت قيمة  $F$  في تحليل التباين ذات دلالة ، أى إذا كان :

$$F = \frac{E^2 / \alpha}{E^2} \leq [F - \alpha, 1 - \nu - E]$$

$$\text{أى } \frac{E^2 (\text{بين الأقسام})}{E^2 (1 - \alpha)} \leq [F - \alpha, 1 - \nu - E]$$

$$\text{أى } E^2 (\text{بين الأقسام}) \leq E^2 (1 - \alpha) [F - \alpha, 1 - \nu - E] \quad (٧)$$

والعدد الذى بالطرف الأيسر من هذه المتباينة هو القيمة الحرجة المطلوبة . ويلاحظ أن هذا العدد أكبر من القيمة الحرجة المناظرة التي كان من الممكن استخدامها في المقارنات القبلية بعد وضع عدد أقسام المقارنة بدلاً من العدد الكلى للأقسام له . والمهدف من كبر هذه القيمة تصعيب تقرير دلالة الاختلاف تعويضاً عن أننا نختار للمقارنة تلك الأقسام التي تسهم إسهاماً كبيراً في دلالة تحليل التباين .

ففي المثال (٨ - ١) وبأخذ  $\alpha = 0.05$  نجد أن :

$$\text{القيمة الحرجة لمجموع المربعات} = 4 \times 0.46 \times 2.58 = 56.35$$

فإذا زادت قيمة مجموع المربعات بين متوسطات قسمين أو أكثر عن هذا العدد أو كانت مساوية له فإنها تكون ذات دلالة عند المستوى  $0.05$  .

### ملاحظة (٦) :

تسمى هذه الطريقة « إجراء الاختبار الآتى لمجموع المربعات » sum of squares simultaneous test procedure وهي إحدى طرق الاختبارات البعدية للمقارنات المتعددة .

### مثال (٨ - ٥) :

اختبر ما تراه يستحق الاختبار في نتائج تجربة المثال (٨ - ١) .

### الحل :

لو رتبنا المتوسطات الناتجة تصاعدياً نجد الآتى :

٥٨ ( جلوكوز + فركتوز ) ، ٥٨,٢ ( فركتوز ) ، ٥٩,٣ ( جلوكوز ) ،  
٦٤,١ ( سكروز ) ، ٧٠,١ ( مراقبة ) . وهذا الترتيب يوحي بعدة مقارنات  
نكتفي منها بما على :

( أولاً ) المقارنة بين المتوسطات الثلاث الأولى ، حيث أنها تبدو قريبة من بعضها ويشك في وجود فرق جوهري بينها .

$$٢٢ \text{ (بين الأقسام الثلاثة)} = \frac{٥٨٠ + ٥٨٢ + ٥٩٣}{٣} - \frac{٥٨٠ + ٥٨٢ + ٥٩٣}{١٠}$$

$$٩,٨ = ١٠٢٦٦٧,٥ - ١٠٢٦٧٧,٣ =$$

بما أن ٩,٨ أصغر من القيمة الحرجة ٥٦,٣٥ نحكم بأن الفروق بين المتوسطات الثلاث ليست ذات دلالة أى يمكن اعتبار أن هذه الأقسام من مجتمعات متساوية المتوسطات ، وذلك عند مستوى الدلالة ٠,٠٥

( ثانياً ) المقارنة بين متوسط السكروز ومتوسط السكريات الثلاثة الأخرى :  
٢٢ ( سكروز ضد السكريات الأخرى ) .

$$= \frac{٥٨٠ + ٥٨٢ + ٥٩٣ + ٦٤١}{٤٠} - \left( \frac{٥٨٠ + ٥٨٢ + ٥٩٣}{٣٠} + \frac{٦٤١}{١٠} \right)$$

$$٢٣٥,٢ = ١٤٣٥٢٠,٤ - ١٠٢٦٦٧,٥ + ٤١٠٨٨,١ =$$

بما أن ٢٣٥,٢ أكبر من القيمة الحرجة ٥٦,٣٥ نحكم بأن الفرق بين السكروز والمستويات الأخرى من السكريات هو فرق جوهري مما يشير إلى أن السكروز يؤخر نمو النبات بدرجة أقل من السكريات الأخرى .

### ملاحظة (٧) :

إذا رغبت في إجراء اختبار بين أى زوج من الأقسام فيمكننا استخدام اختبار ت للمقارنة بين متوسطى مجتمعين متتليين - راجع البند (٦ - ٦ - ٣) مع أخذ  $\bar{C} = \bar{C}'$  لتقدير التباين أى نستخدم الإحصاء :

$$\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \cdot \epsilon$$

بدرجات حرية (n - k) وهى درجة الحرية لتقدير التباين داخل الأقسام في تحليل التباين ، إلا أن الطريقة سابقة الذكر أكثر عمومية ، فهى صالحة للتطبيق مهما كان عدد أقسام المقارنة .

كما يمكننا إيجاد حدى ثقة بدرجة (α - ١) للفرق بين المتوسطين من الصيغة .

$$(٨) \quad (\bar{y}_1 - \bar{y}_2) \pm \epsilon \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{ت} \quad (\alpha - ١)$$

أما حدا الثقة بدرجة (α - ١) لم توسط أى قسم  $\bar{y}_i$  فهما :

$$(٩) \quad \bar{y}_i \pm \epsilon \sqrt{\frac{1}{n_i}} \quad \text{ت} \quad (\alpha - ١)$$

ملاحظة (٨) : القيمة الحرجة للفرق بين متوسطى عيتين

#### CRITICAL DIFFERENCE (C.D.)

إذا رغبت في إجراء عدة اختبارات بعدية بين أزواج من المتوسطات فيمكن بنفس فكرة القيمة الحرجة لمجموع المربعات إيجاد القيمة الحرجة للفرق بين أى زوج  $\bar{y}_1$  ،  $\bar{y}_2$  من المتوسطات كالآتى ( على فرض تساوى عدد الوحدات في كل مجموعة ) :

يكون الفرق بين متوسطين  $\bar{y}_1$  ،  $\bar{y}_2$  ذا دلالة

$$\text{إذا كان } \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t_{\alpha} (n - 2)$$

$$\text{أى إذا كان } |\bar{y}_1 - \bar{y}_2| \leq \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{\alpha} (n - 2)$$

والعدد الذى بالطرف الأيسر من هذه المتباينة هو القيمة الحرجة المطلوبة .

ففى المثال (٨ - ١) وبأخذ  $\alpha = 0.05$  نجد أن :

$$\begin{aligned} \text{القيمة الحرجة بين متوسطى قسمين} &= \sqrt{0.467 \times 0.2 \times 2.01} \\ &= 2.1005 \end{aligned}$$

إذا زاد الفرق  $|\bar{y}_1 - \bar{y}_2|$  عن العدد ٢,١٠٠٥ أو كان مساوياً له فإن هذا الفرق يكون ذا دلالة عند المستوى ٠,٠٥ وإلا فلا دلالة له عند هذا المستوى .

فمثلاً : الفرق بين متوسطى قسمى الجلوكوز والفركتوز  $= 58.2 - 59.3 = 1.1$

وهذا الفرق أصغر من القيمة الحرجة ٢,١ فهو غير ذى دلالة عند المستوى

٠,٠٥ . أما الفرق بين متوسطى قسمى السكروز والمراقبة وهو  $70.1 - 64.1 = 6$

فأكبر من القيمة الحرجة ٢,١ فهو فرق جوهري . ونصل إلى نفس هذه الاستنتاجات إذا استخدمنا طريقة القيمة الحرجة لمجموع المربعات .

### ملاحظة (٥) : التقسيم الأحادى

#### ONE—WAY CLASSIFICATION

إن التقسيم الناتج عن التجارب ذوات العامل الواحد التى نوقشت فى البنود

السابقة يندرج تحت ما يسمى بالتقسيم الأحادى أو التقسيم من ناحية واحدة .  
وهذا التقسيم يتناول حالتين :

(أ) الحالة السابق دراستها ومثل لها بالمثل ( ٨ - ١ ) وتوابعه ، حيث يكون لدينا مجتمع واحد ونريد اختبار تأثيره من المعالجات ( مستويات عامل التقسيم ) على متغير ما متعلق بهذا المجتمع . فى هذه الحالة نسحب من هذا المجتمع عينة عشوائية ثم نقسمها عشوائيا إلى ك من المجموعات غير المتداخلة ونعطى لكل منها معالجة مختلفة ، ثم ندرس الاستجابات فى هذه المجموعات .

(ب) الحالة التى يكون لدينا فيها ك من المجتمعات لكل منها خاصية متميزة ويراد دراسة تأثير هذه الخواص على متغير ما . هنا نسحب عينة عشوائية من كل مجتمع ونعطى لكل منها نفس المعالجة ثم ندرس الاستجابات . ومن أمثلة هذه الحالة المسائل ٢ ، ٣ ، ٥ ، ٦ من التمارين ( ٨ - ١ ) الآتية . ففى المسألة (٢) لدينا ٤ مجتمعات هى مجتمعات الأنواع الأربعة من الأرناب ، وفى المسألة (٣) لدينا ٣ مجتمعات هى المجتمع الذى تزرع فيه البذور فى أول مايو والمجتمع الذى تزرع فيه البذور فى ١٥ مايو ثم مجتمع ٢٩ مايو .

فى الحالة (ب) نفترض أن المجتمعات التى سحبت منها العينات هى مجتمعات معتدلة لها نفس التباين ، ونستخدم نفس النموذج الإحصائى ، وبهذا لا يختلف التحليل الإحصائى نظريا ولا حسابيا عن الحالة (أ) .

## تمارين ( ٨ - ١ )

### ( الفروض العشوائية والاعتدالية متوفرة )

(١) الجدول الآتى يبين عينة من قيم شدة المقاومة لمعدن معين ( بعد طرح ١٠٠ من كل منها ) وقد حصلنا على هذه القيم من ٤ أشرطة من المعدن قيس كل منها

عند ٣ نقط مختلفة ( الركن - الوسط - الحافة ) . هل متوسط شدة المقاومة واحد عند جميع نقط الشريحة ؟

الحافة	الوسط	الركن
٤٢	٤٠	٣٧
٤٠	٣٩	٤٢
٣٣	١٧	٢٨
٤١	٣٧	٣٧

(٢) القيم الآتية هي أطوال أذنان نوع معين من يرقات القراد في عينات من ٤ أنواع من الأرناب ( مقيسة بالميكرون ) . هل هناك فروق جوهرية بين متوسطات الأطوال في هذه الأنواع ؟

(١)	(٢)	(٣)	(٤)
٣٨٠	٣٥٠	٣٥٤	٣٧٦
٣٧٦	٣٥٦	٣٦٠	٣٤٤
٣٦٠	٣٥٨	٣٦٢	٣٤٢
٣٦٨	٣٧٦	٣٥٢	٣٧٢
٣٧٢	٣٣٨	٣٦٦	٣٧٤
٣٦٦	٣٤٢	٣٧٢	٣٦٠
٣٧٤	٣٦٦	٣٦٢	
٣٨٢	٣٥٠	٣٤٤	
	٣٤٤	٣٤٢	
	٣٦٤	٣٥٨	
		٣٥١	
		٣٤٨	
		٣٤٨	

اطرح ٣٠٠ أو أى عدد مناسب

(٣) هل تاريخ زراعة القطن يؤثر في وزن المحصول الناتج من البنور ؟  
الآتي هي الأوزان بالكيلوجرامات الناتجة من ٤ حقول قسم كل منها إلى ٣ أحواض :

٢٩ مايو	١٥ مايو	١ مايو
١,٩٩	٣,٨٦	٣,٣٥
٢,٨٩	٢,٧١	١,٤٩
١,٦٨	٢,١٨	٢,٤٤
٢,١٣	١,٩٥	٢,٤٤

(٤) قسم ٢٨ أرنياً عشوائياً إلى ٤ أقسام متكافئة وأعطى لكل قسم معالجة ما (مثلاً : دواء - نظام تغذية - بيئة ..) والجدول الآتي يعطي مستوى السكر في الدم الناتج من هذه المعالجات . هل هناك دليل على وجود فروق بين تأثيرات هذه المعالجات على مستوى السكر في الدم ؟

#### المعالجات

(٤)	(٣)	(٢)	(١)
٩	٣٥	٢٧	١٧
٨	٢٢	٣٦	١٦
١٧	٣٥	٢١	٢٨
١٨	٢٨	١٣	٤
١	٣١	٤٥	٢١
٣٤	٣٤	٢٣	صفر
١٣	٤٠	١٣	٢٣



(٥) في تجربة صناعية كان أحد المهندسين مهتماً بكيفية تغير متوسط امتصاص الرطوبة بين خمس مجموعات مختلفة من الخرسانة ، واستخدم لذلك ٦ عينات لكل مجموعة تعرضت للرطوبة لمدة ٤٨ ساعة فجاءت البيانات كما يلي .

١	٢	٣	٤	٥
٥٥١	٥٩٥	٦٣٩	٤١٧	٥٦٣
٤٥٧	٥٨٠	٦١٥	٤٤٩	٦٣١
٤٥٠	٥٠٨	٥١١	٥١٧	٥٢٢
٧٣١	٥٨٣	٥٧٣	٤٣٨	٦١٣
٤٩٩	٦٣٣	٦٤٨	٤١٥	٥٥٦
٦٣٢	٥١٧	٦٧٧	٥٥٥	٦٧٩
اطرح ٤٥٠ أو ٥٠٠ أو أى عدد مناسب .				

اختبر أن  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$  عند المستوى ٠,٠٥ .

(٦) في تجربة ما اختبرت سلالتان من *Drosophila Melanogaster* إحداهما لدورة يرقات قصيرة (short larval period) والأخرى لدورة طويلة ، كما أخذت سلالة كمجموعة مراقبة . وفي الجيل ٤٢ لخصت البيانات عن طول الدورة بالساعات فيما يلي :

السلالة		
دورة قصيرة	دورة طويلة	مراقبة
٨٠	٣٣	٦٩
٨٠٧٠	٣٦٤٠	٧٢٩١
١٩٩٤٦٥٠		

أولاً : أجر تحليل التباين وفسره .

ثانياً : أجر المقارنة القبلية بين الدورتين القصيرة والطويلة معاً ضد المراقبة .

ثالثاً : أجر المقارنة البعدية بين كل من أزواج المتوسطات الثلاثة .

رابعاً : أوجد ضرة ثقة بدرجة ٩٥٪ لكل من هذه المتوسطات .

(٧) في أحد الأبحاث الطبية عن علاج جحوظ العين الناتج من تسمم الغدة الدرقية اختبرت عينة عشوائية من ٣٠ ضفدعة وقسمت عشوائياً إلى ثلاث مجموعات متساوية العدد تلقت إحداها علاجاً دوائياً وتلقت أخرى علاجاً جراحياً ( استئصال الغدة النخامية ) وتركزت المجموعة الثالثة دون علاج ( مجموعة مراقبة ) ثم قيس العيون بالمقياس الأفقي للعين فتتجت القيم الآتية بالملليمترات .

مجموعة العلاج الدوائي : ١,٧ ١,٤ ١,٨ ١,٩ ١,٨ ١,٣ ١,٦ ١,٥ ١,٧

مجموعة العلاج الجراحي : ١,٤ ١,٥ ١,٣ ١,٢ ١,٨ ١,٩ ١,٩ ١,٧ ١,٢

مجموعة المراقبة : ١,٩ ٢ ١,٨ ٢ ١,٩ ٢ ٢ ١,٧ ١,٨

( أ ) اختبر تأثير مجموعتي العلاج ضد مجموعة المراقبة .

( ب ) اختبر ما إذا كان هناك فرق جوهري بين نوعي العلاج .

( خذ  $\alpha = 0,01$  )

## ( ٨ - ٦ ) التجارب ذات العاملين :

### TWO-FACTOR EXPERIMENTS

اعتبر عينة حجمها  $n$  مأخوذة من متغير مختل  $M$  وسطه الحسابي  $\mu$  وتباينه  $\sigma^2$  وافرض أن وحدات هذه العينة قد تعرضت لتأثير عاملين لأحدهما  $k$  من المستويات وللثاني  $h$  من المستويات . المطلوب اختبار ما إذا كان المجتمع الذي أخذت منه العينة متجانساً بالنسبة لكل من هذين العاملين على حدة .

في هذه الحال علينا أن نختار بين طريقتين للتحليل ، ويتوقف هذا الاختيار على ما إذا كان العاملان مستقلين أو كانا يتفاعلان معاً . والمقصود بكلمة التفاعل هنا هو تأثير أى من العاملين على الآخر . وسنبداً بالحالة الأبسط التي سنفترض فيها عدم وجود تفاعل بين العاملين .

### ( ٨ - ٦ - ١ ) حالة عاملين لا يتفاعلان :

على أساس توفر عوامل الاعتدالية والعشوائية وخطية النموذج وعلى فرض عدم وجود تفاعل بين عاملى التجريب ، يكون تحليل التباين ما هو إلا امتداد لتحليل التباين للتجارب ذوات العامل الواحد ولا يزيد عنه إلا بخطوة منطقية واحدة .  
توضع وحدات العينة في ك من الأعمدة تناظر مستويات العامل الأول ، هـ من الصفوف تناظر مستويات العامل الثاني ، فتخضع أى وحدة تجريبية  $\mu_{gh}$  للنموذج الآتي :

$$\mu_{gh} = \mu + \alpha_g + \beta_h + \gamma_{gh} \quad (١٠)$$

حيث  $\mu$  ،  $\alpha_g$  ،  $\beta_h$  ،  $\gamma_{gh}$  تحمل نفس المعاني السابقة في النموذج (٢) ، وحيث  $\beta_h$  تعبر عن متوسط أثر المستوى  $h$  للعامل الثاني .

وكنتيجة مباشرة لذلك ، لا يتحلل مجموع المربعات الكلى إلى مركبتين كما هو الحال في المتساوية (٣) بل إلى ثلاث مركبات . ومن الطبيعي أن تظل المركبة الأولى التي تعبر عن الاختلاف بين مستويات العامل الأول ( الأعمدة ) كما هي ، أما المركبة الثانية في المتساوية (٣) فتفصل إلى مركبتين لإحداهما تعبر عن الاختلاف بين مستويات العامل الثاني ( الصفوف ) والأخرى تعطى الاختلاف الذى يبقى من الاختلاف الكلى بعد استبعاد أثر كل من العاملين . هذا الاختلاف يعزى إلى عدة أسباب نضمها تحت كلمة « الخطأ » ، منها الخطأ العشوائى أو خطأ التجريب ومنها الخطأ الناشئ عن إهمال التفاعل بين العاملين إذا كان هناك تفاعل .

يمكن جبرياً أن تثبت المتطابقة الآتية :

$$\frac{1}{2} \frac{d}{dt} (s_1 - s_2) = \frac{1}{2} \frac{d}{dt} (s_1 - s_2) + \frac{1}{2} \frac{d}{dt} (s_1 - s_2)$$

$$(9) \quad \frac{1}{2} \frac{d}{dt} (s_1 - s_2) = \frac{1}{2} \frac{d}{dt} (s_1 - s_2) + \frac{1}{2} \frac{d}{dt} (s_1 - s_2)$$

وهذه المتطابقة تترجم لفظياً كالآتي :

$$22 = (22) \text{ (بين الأعمدة) } + 22 \text{ (بين الصفوف) } + 22 \text{ (الخطأ) }$$

$$\text{حيث } 1 = 1, 1 = 1, 1 = 1$$

$$(10) \quad (1 - 1) = (1 - 1) - (1 - 1) = 1$$

وإذن :

$$1 = 1 + 1 + 1$$

إن هذه المركبات الثلاث تعطي ثلاثة تقديرات مستقلة لتباين المجتمع  $\sigma^2$  وهي :

$$1 = 1 = 22 \text{ (بين الأعمدة) } / (1 - 1), 1 = 22 \text{ (بين الصفوف) } / (1 - 1),$$

$$1 = 22 \text{ (الخطأ) } / (1 - 1 - 1 + 1)$$

ولا يبق إلا استخدام نسبة التباين  $\frac{1}{2} \frac{d}{dt}$  لاختبار صحة الفرض الصفري عن

تساوي متوسطات مستويات العامل الأول . واستخدام نسبة التباين  $\frac{1}{2} \frac{d}{dt}$  لاختبار

صحة الفرض الصفري عن تساوي متوسطات مستويات العامل الثاني .

## مثال (٨ - ٦) :

قسمت ١٢ بقرة إلى هـ = ٤ من المجموعات بكل منها ٣ بقرات بحسب الوزن عند بدء التجربة . أعطى للأبقار الثلاث بكل مجموعة نوع مختلف من الغذاء . وبعد فترة من الزمن قيست الزيادات في أوزان الأبقار ورصدت بالجدول (٨ - ٦) الآتي . لدينا عاملان الأول هو عامل الغذاء وله هـ = ٣ مستويات والثاني هو عامل الوزن الابتدائي للأبقار وله هـ = ٤ مستويات . المطلوب بحث تأثير كل من هذين العاملين على الزيادة في وزن الأبقار عند مستوى الدلالة ٠,٠٥.

جدول (٨ - ٦)

الوزن الابتدائي للأبقار	نوع الغذاء			٢	٣
	١	٢	٣		
١	٧,٠	١٤,٠	٨,٥	٣	٢٩,٥
٢	١٦,٠	١٥,٥	١٦,٥	٣	٤٨,٠
٣	١٠,٥	١٥,٠	٩,٥	٣	٣٥,٠
٤	١٣,٥	٢١,٠	١٣,٥	٣	٤٨,٠
٥	٤	٤	٤	١٢ = هـ	
٦	٤٧,٠	٦٥,٥	٤٨,٠	١٦٠,٥ = ٢	

## الحل :

لدينا اثنان من الفروض الصفرية : لا الوزن الابتدائي ولا نوع الغذاء له تأثير في زيادة أوزان الأبقار .

$$22 \text{ (الكل)} = 7 + \dots + 14 + \dots + 18 + \dots + 13 + \dots - \frac{160,0}{12}$$

$$= 2146,6875 - 2316,75 =$$

$$170,0625 = \text{بدرجات حرية } 11$$

$$22 \text{ (بين الأعمدة)} = \frac{47 + 60 + 48}{4} - \frac{160,0}{12}$$

$$= 2146,6875 - 2200,8125 =$$

$$54,125 = \text{بدرجات حرية } 2$$

$$22 \text{ (بين الصفوف)} = \frac{29,0 + 48 + 30 + 48}{3} - \frac{160,0}{12}$$

$$= 2146,6875 - 2234,4166 =$$

$$87,7291 = \text{بدرجات حرية } 3$$

$$22 \text{ (الخطأ)} = 170,0625 - (54,125 + 87,7291) =$$

$$28,2084 = \text{بدرجات حرية } 6$$

وينشأ جدول التباين الآتي .

الجدول (٨ - ٧)

مصدر التباين	المجموع للمربعات	درجات الحرية	تقدير التباين	د.ف
بين الأعمدة (نوع الطعام)	54,1250	2 = 1 - 0	$27,0625 = \frac{7}{2}$	5,756
بين الصفوف (الوزن الإجمالي)	87,7291	3 = 1 - 0	$29,2430 = \frac{7}{2}$	9,220
خطأ التجريب	28,2084	6 = (1-2)(1-0)	$4,7014 = \frac{7}{2}$	
المجموع	170,0625	11 = 1 - 0		

من الجدول نجد أن  $F_{[6, 27], 0.05} = 5.14$  وهذه أصغر من  $5.756$  وإذن نرفض الفرض الصفري الأول عند المستوى  $0.05$  , ونقرر أن اختلاف نوع الغذاء يؤدي إلى الاختلاف في الزيادة في أوزان الأبقار .

كذلك  $F_{[6, 27], 0.01} = 4.76$  وهذه أصغر من  $6.220$  وإذن نرفض الفرض الصفري الثاني عند المستوى  $0.01$  , ونقرر أن اختلاف الأوزان الابتدائية للأبقار له تأثير في الزيادة في أوزانها .

## تمارين (٨ - ٢)

(١) أجريت تجربة زراعية لاختبار تأثير اختلاف التربة (٥ قطع من الأرض) واختلاف نوع القمح (٧ سلالات) على محصول الحبوب . وقد قسمت كل قطعة أرض عشوائياً إلى ٧ أحواض وزعت عليها السلالات السبع عشوائياً فنتجت البيانات الآتية التي تسجل المقادير الناتجة للمحصول بالكيلو .

السلالة	القطع							مجموع
	(١)	(٢)	(٣)	(٤)	(٥)	(٦)	(٧)	
(١)	١٦	١٥	١٧	١٤	١٤	١٥	١٣	١٠٤
(٢)	١٢	٩	١٥	١٠	١٠	١١	١١	٧٨
(٣)	١٣	١٣	١٤	١٥	١٢	١٣	١٠	٩٠
(٤)	١٥	١٤	١٩	١٧	١٣	١٨	١٦	١١٢
(٥)	١١	١٠	١٢	١٠	١١	١٢	١٢	٧٨
مجموع	٦٧	٦١	٧٧	٦٦	٦٠	٦٩	٦٢	٤٦٢

$$6314 = \sum_{i=1}^5 \sum_{j=1}^7 x_{ij}^2$$

أولاً : ابحث دلالة تأثير كل من عاملى التربة ونوع القمح .  
 ثانياً : ابحث ما إذا كان الفرق بين السلالتين (٥) ، (٦) ذا دلالة عند المستوى ٠,٠٥ .  
 ثالثاً : ابحث ما إذا كانت تربة القطعتين ٢ ، ٥ ( معاً ) تختلف عن تربة القطعتين ١ ، ٤ ( معاً ) .

( استخدام مستوى الدلالة ٠,٠١ ) .

(٢) الآتي هى مقادير الكلوسترول ( بالمليجرام فى العبوة ) التى وجدتها ٤ معامل فى عبوات لثلاثة أنواع متشابهة من الغذاء وزن كل منها ٦ أوقيات .

المعامل	الغذاء			
	(١)	(٢)	(٣)	(٤)
(أ)	٣,٧	٢,٨	٣,١	٣,٤
(ب)	٣,١	٢,٦	٢,٧	٣,٠
(ج)	٣,٥	٣,٤	٣,٠	٣,٣

اختبر عند مستوى الدلالة ٠,٠٥ ما إذا كانت :

( أولاً ) متوسطات الكلوسترول فى الأنواع الثلاثة من الغذاء متساوية .  
 ( ثانياً ) المتوسطات التى حصلت عليها المعامل الأربعة متساوية .  
 ( ٣ ) الأعداد الآتية هى درجات الحرارة ( مقاسة بالستجراى ) لمياه إحدى البحيرات فى أربعة أيام متتالية من صيف ١٩٥٢ م ( الساعة الثانية بعد الظهر ) ، وقد أخذت هذه الدرجات على ١٠ أعماق مختلفة ( مقاسة بالمتر ) . ابحث دلالة كل من عاملى اليوم والعمق .



الأعماق	٢٩ يوليو	٣٠ يوليو	٣١ يوليو	١ أغسطس	م
٠	٢٣,٨	٢٤,٠	٢٤,٦	٢٤,٨	٩٧,٢
١	٢٢,٦	٢٢,٤	٢٢,٩	٢٣,٢	٩١,١
٢	٢٢,٢	٢٢,١	٢٢,١	٢٢,٢	٨٨,٦
٣	٢١,٢	٢١,٨	٢١,٠	٢١,٢	٨٥,٢
٤	١٨,٤	١٩,٣	١٩,٠	١٨,٨	٧٥,٥
٥	١٣,٥	١٤,٤	١٤,٢	١٣,٨	٥٥,٩
٦	٩,٨	٩,٩	١٠,٤	٩,٦	٣٩,٧
٩	٦,٠	٦,٠	٦,٣	٦,٣	٢٤,٦
١٢,٥	٥,٨	٥,٩	٦,٠	٥,٨	٢٣,٥
١٥,٥	٥,٦	٥,٦	٥,٥	٥,٦	٢٢,٣
م	١٤٨,٩	١٥١,٤	١٥٢,٠	١٥١,٣	٦٠٣,٦

$$\bar{y} = 11230.78$$

#### (٨ - ٦ - ٢) حالة عاملين يتفاعلان :

في التجارب ذوات العاملين ، إذا كان هناك شك في وجود تفاعل بين العاملين أى في تأثير كل منهما جزئياً بالآخر ، فإن النموذج (١٠) لا يكون مناسباً لأنه لا يفسح مكاناً لتأثير هذا التفاعل . وعلى فرض توفر شروط الاعتدالية والعشوائية وخطية النموذج فإن النموذج المناسب يأخذ الصورة الآتية :

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad (11)$$

حيث  $(\beta\alpha)$  تدبر عن تأثير الاختلاف الناشئ من تفاعل العاملين . ولكي نوجد مقياساً لهذا الاختلاف لا مفر من تكرير التجربة برمتها مرة واحدة على الأقل وذلك لأنه في التجربة الواحدة يؤثر المستوى ٣ مثلاً من العامل الأول مع المستوى ٢ مثلاً من العامل الثاني على وحدة واحدة فقط هي ٣٣ من وحدات التجريب ، وبالمثل بالنسبة لأزواج المستويات الأخرى ، ولا يكون هناك مجال حينئذ لإيجاد الاختلاف الناشئ عن تفاعل العاملين إلا إذا كان هناك أكثر من وحدة تتعرض لتأثير كل من أزواج هذه المستويات ، أى إلا إذا كررت التجربة مرة واحدة على الأقل . وتجمع نتيجة التجريبتين أو التجارب فيما يسمى بالخلايا كما في المثال (٨ - ٧) الآتي .

ومن الناحية الحسابية نوجد كلا من ٢٢ ( الكلى ) ، ٢٢ ( بين الأعمدة ) ، ٢٢ ( بين الصفوف ) من واقع القيم الناتجة عن التجريب كما في البند السابق . أما بالنسبة للاختلافين الباقيين فتحسبهما من مجاميع الخلايا كالآتي :

نفرض أننا كررنا التجربة برمتها ١ من المرات فيكون كل زوج من أزواج المستويات قد أثر في ١ من وحدات التجريب . سنسمى مجموع قيم كل من هذه الوحدات « مجموع الخلية » ونرمز له بالرمز  $\mu_{ij}$  وإذن :

$$\mu_{ij} = \frac{\mu_{i.}}{j} - \frac{\mu_{.j}}{i} \quad (١٢)$$

بدرجات حرية عددها (ك - هـ - ١) حيث ك عدد الأعمدة ، هـ عدد الصفوف . إن هذا الاختلاف هو مجموع الاختلافات الناشئة عن العاملين بكل أنواعها وإذن :

$$\mu_{ij} = (\text{تفاعل العاملين}) = \mu_{ij} - (\text{بين الخلايا}) - (\text{بين الأعمدة}) -$$

$$(\text{بين الصفوف}) - \text{بدرجات حرية عددها}$$

$$(\text{ك} - \text{هـ} - ١) - (\text{ك} - ١) - (\text{هـ} - ١) = \text{ك} - \text{هـ} - ١ - ١ + ١ =$$

$$= (\text{ك} - ١) - (\text{هـ} - ١)$$

أما الاختلاف الذى نضعه تحت كلمة « خطأ » فهو الاختلاف المتبقي من الاختلاف الكلى بعد استبعاد جملة الاختلاف الناشء من العاملين . أى أن :

$$\begin{aligned} ٢٢ ( \text{الخطأ} ) &= ٢٢ ( \text{الكلى} ) - ٢٢ ( \text{بين الخلايا} ) . \\ \text{بدرجات حرية عندها } (١ - ١) - (١ - ١) &= ١ - ١ = ٠ \end{aligned}$$

مثال (٨ - ٧) :

الجدول (٨ - ٨) الآتي يعطى نتائج تجربة أجريت مرتين (بشكل مستقل) للدراسة تأثير كل من عاملى شدة الضوء ودرجة الحرارة على معدل نمو أحد النباتات . الأعداد ١٧ ، ٦ ، ٣ ، ٢ ، ٠ ، ٠ ، ٠ ، ٨ ، ٤ ، ٥ ، ٢ هي نتائج التجربة الأولى والأعداد ٩ ، ٨ ، ٤ ، ٢ ، ٠ ، ٠ ، ٠ ، ١٠ ، ٥ ، ٤ ، ٢ هي نتائج التجربة الثانية . أما الأعداد التى بين قوسين فهى مجاميع الخلايا . أجر تحليل التباين ثم أجب عن التساؤلات القليلة الآتية :

( أولاً ) هل متوسط تأثير درجة الحرارة ٨ يساوى متوسط تأثير درجة الحرارة ؟ ٢٨

( ثانياً ) هل متوسط تأثير درجة الحرارة ١٤ يساوى متوسط تأثير درجة الحرارة ؟ ٢١

( ثالثاً ) هل متوسط تأثير درجتى الحرارة ٨ ، ٢٨ معا يساوى متوسط تأثير درجتى الحرارة ١٤ ، ٢١ معا ؟

الحل :

$$\begin{aligned} ٢٢ ( \text{الكلى} ) &= \frac{٢٢}{١} - \frac{٢٢}{١} = ٢٢ - ٢٢ = ٠ \\ ١٧ + ٢ + ٣ + ٦ + ٨ + \dots + ١٠ + ٥ + ٤ + ٢ + ٢ &= ١١٧٦ - ١١٧٤ = ٢ \\ ٣١ - ٣٢ &= -١ \end{aligned}$$

المجدول (أ - أ)

درجة الحرارة	(١)	(٢)	(٣)	(٤)	نوع	كم
أ	(٢٦) ٩، ١٧	(٢٨) ١٥، ١٣	(٢٩) ١٦، ١٣	(١٨) ١٠، ٨	أ	١٠١
١٤	(١٤) ٨، ٦	(١١) ٥، ٦	(١٠) ٤، ٦	(٩) ٥، ٤	أ	٤٤
٢١	(٧) ٤، ٣	(٨) ٥، ٣	(٧) ٤، ٣	(٩) ٤، ٥	أ	٣١
٢٨	(٤) ٣، ٢	(٥) ٣، ٢	(٥) ٣، ٢	(٤) ٣، ٢	أ	١٨
نوع	أ	أ	أ	أ	٣٢ = ٥	
كم	٥١	٥٢	٥١	٤٠	١٩٤ = ٢	

$$٢٢ \text{ (بين الأعمدة) } = \frac{\sum_{i=1}^n \frac{1}{n_i} - \frac{1}{n}}{\frac{1}{n}} =$$

$$١١٧٦ - \frac{1}{٨} (10 + 44 + 31 + 18) =$$

$$١١٧٦ - ١١٨٨,٢٥ =$$

حيث  $\frac{1}{n} = ٣$

$$١٢,٢٥ =$$

$$٢٢ \text{ (بين الصفوف) } = \frac{\sum_{j=1}^m \frac{1}{m_j} - \frac{1}{m}}{\frac{1}{m}} =$$

$$١١٧٦ - \frac{1}{٨} (18 + 31 + 44 + 10) =$$

حيث  $\frac{1}{n} = ٣$

$$٥٠١,٧٥ = ١١٧٦ - ١٦٧٧,٧٥ =$$

$${}^{22} \text{ (بين الخلايا) } = \frac{{}^2 \text{ م}}{1} - \frac{{}^2 \text{ ن}}{2}$$

$$1176 - \frac{{}^2 4 + {}^2 9 + \dots + {}^2 14 + {}^2 27}{2} =$$

$$1176 - 1724 =$$

$$10 = 1 - 4 \times 4 = \text{حيث } \text{م} = 048$$

$$(0.1, 70 + 12, 20) - 048 = \text{م (تفاعل)}$$

$$9 = (1 - 4)(1 - 4) = \text{حيث } \text{م} = 34$$

$$048 - 098 = \text{م (الخطأ) ،}$$

$$16 = 10 - 31 = \text{حيث } \text{م} = 00$$

جدول (8 - 9)

ن	تقريب العين	درجات الحرية	مجموع المربعات	مصدر العين
1, 3.4	4, 0.8	3	12, 20	بين الأصعدة (شدة الإصابة)
02, 430	167, 20	3	0.1, 70	بين الصفوف (درجة الحرارة)
1, 2.4	3, 77	9	34	تفاعل (الإصابة × الحرارة)
	3, 13	16	00	داخل الأقسام (الخطأ)
		31	098	المجموع

ف... (11, 3) تقع بين 3, 10 ، 3, 29 ، ف... (16, 3) تقع بين 4, 94 ، 0, 42

ف... (11, 9) تقع بين 39 ، 209

وعلى ذلك فإن :

- (١) الاختلافات بين مستويات شدة الإضاءة ليست ذات دلالة عند المستوى ٠,٠٥  
( أى يمكن اعتبار هذه المستويات متكافئة ) .
- (٢) الاختلافات بين مستويات درجة الحرارة ذات دلالة عالية .
- (٣) التفاعل بين عامل الإضاءة ودرجة الحرارة ليس ذو دلالة عند المستوى ٠,٠٥  
نحجب الآن عن التساؤلات الثلاثة المطروحة .  
( أولا )

$$٢٢ \text{ (درجة الحرارة ٨ ضد درجة الحرارة ٢٨) } = \frac{119}{16} - \frac{118}{8} + \frac{110}{8} = ٤٣٠,٥٦٢٥$$

$$١٣٧,٥٦ = \frac{٤٣٠,٥٦٢٥}{٣,١٣} = \text{ في }$$

بما أن  $F_{[16,13],0.01} = ٨,٦٨ > ١٣٧,٥٦$  نرفض الفرض الصفري عن تساوى تأثير  
درجة الحرارة ٨ وتأثير درجة الحرارة ٢٨ ونستنتج أن للدرجة الحرارة ٨ تأثير أكبر ،  
وذلك عند مستوى الدلالة ٠,٠١  
( ثانيا )

$$٢٢ \text{ (بين مستوى درجتى الحرارة ١٤ ، ٢١) } = \frac{75}{16} - \frac{73}{8} + \frac{44}{8} = ١٠,٥٦٢٥$$

$$٣,٣٧٥ = \frac{١٠,٥٦٢٥}{٣,١٣} = \text{ في }$$

بما أن  $F_{[16,13],0.05} = ٤,٥٤ < ٣,٣٧٥$  نقبل الفرض الصفري عن تساوى تأثير  
درجتى الحرارة ١٤ ، ٢١ .

( ثالثاً )

$$F_{(22, 16)} = \frac{(198 + 101)}{16} + \frac{(31 + 44)}{16} - \frac{194}{32} = 60,5$$

$$F_{(22, 16)} = \frac{60,5}{3,13} = 19,329$$

نرفض الفرض الصفري عند مستوى الدلالة 0,01

### ملاحظة (٩) :

إن هذه المقارنات الثلاث مستقلة ويمكن التأكد من ذلك باستخدام معيار استقلال مقارنتين الذى سقدمه بالبند (٨ - ١١ - ١) . ولذلك فإن مجموع مجاميع المربعات لهذه المقارنات وهو  $430,5625 + 10,5625 + 60,5 = 501,625$  يساوى مجموع المربعات بين الصفوف ( درجات الحرارة ) ، أى أن هذا المجموع قد تحلل إلى ٣ مركبات مستقلة ، كما تحللت درجات حريرته الثلاثة إلى ثلاث درجات حرية منفصلة .

### ملاحظة (١٠) :

إذا كنا قد وجدنا تفاعلاً ذا دلالة بين شدة الإضاءة ودرجة الحرارة فإن الخطوة التالية تكون حينئذ تقسيم البيانات إلى جداول منفصلة لكل مستوى من مستويات شدة الإضاءة وتحليل كل جدول لاختبار تأثير مستويات درجة الحرارة ، وقد نجد أنه في مستوى ما من مستويات الإضاءة تكون درجة الحرارة ذات تأثير جوهري بينما في مستوى آخر لا يكون هذا التأثير جوهرياً . وبالمثل يمكن تجزئ البيانات لكل من مستويات درجة الحرارة لاختبار تأثير مستوى الإضاءة عند كل درجة حرارة .

### تقارين (٨ - ٣)

في دراسة استهلاك الأكسجين لسلالتين من الحيوانات الصدفية عند ثلاثة تركيزات لماء البحر أخذت ٤ قراءات لكل مركب من السلالة وتركيز الماء (الملوحة) وسجلت القراءات بمقياس معين في الجدول الآتي :

المجموع	السلالة		الملوحة
	(٢)	(١)	
	٦,١٤	٧,١٦	%١٠٠
	٣,٨٦	٦,٧٨	
	١٠,٤٠	١٣,٦٠	
	٥,٤٩	٨,٩٤	
٦٢,٣٦	٢٥,٨٩ = $\Sigma$	٣٦,٤٧ = $\Sigma$	%٧٥
	٤,٤٧	٥,٢٠	
	٩,٩٠	٥,٢٠	
	٥,٧٥	٧,١٨	
٥٥,٨٧	١١,٨٠	٦,٣٧	%٥٠
	٣١,٩٢ = $\Sigma$	٢٣,٩٥ = $\Sigma$	
	٩,٦٣	١١,١١	
	٦,٣٨	٩,٧٤	
٩٣,٣٠	١٣,٤٠	٢٨,٨٠	%٥٠
	١٤,٥٠	٩,٧٤	
	٤٣,٩١ = $\Sigma$	٤٩,٣٩ = $\Sigma$	
٢١١,٥٣	١٠١,٧٢	١٠٩,٨١	المجموع



- (١) اختبر الفرض أن استهلاك الأوكسجين لا يختلف باختلاف السلالة .
  - (٢) اختبر الفرض أن استهلاك الأوكسجين لا يختلف باختلاف الملوحة .
  - (٣) اختبر تأثير تفاعل السلالة والملوحة على استهلاك الأوكسجين .
- ( خذ  $\alpha = 0.05$  )

## ٨ - ٧) المقارنات التزاوجية : PAIRED COMPARISONS

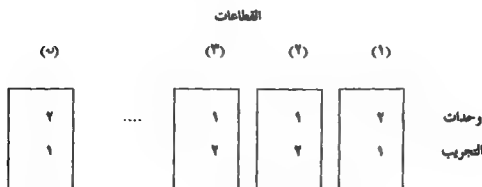
المفروض في التجارب ذوات العامل الواحد أن تكون وحدات التجريب في مختلف أقسام المعالجة متجانسة بقدر الإمكان ، وذلك لكي تكون الفروق في استجابات هذه الوحدات لمستويات العامل راجعة فقط للفروق بين هذه المستويات . أما إذا كانت الوحدات غير متجانسة فإن هذا يتيح الفرصة لاختلافات عشوائية قد تكون من الكبر بحيث تؤدي إلى اختفاء الفروق الحقيقية في تأثيرات تلك المستويات كما سنرى في المثال (٨ - ٨) القادم .

غير أن مطلب التجانس هذا قد يفرض قيوداً عنيفة على عدد الوحدات التي تحتاجها الدراسة ، فمثلاً لمقارنة نوعين من المسكنات لمرض ما المفروض أن يكون المرضى من نفس الجنس ومن نفس العمر والظروف الصحية وشدة الألم الناتج عن المرض ... ومن الواضح أنه من الصعب عملياً الحصول على عدد كاف من المرضى الذين يشتركون في هذه الخواص . وحتى إذا أمكن الحصول على مجموعة بهذه الأوصاف فإن دراسة مثل هذه المجموعة الخاصة تكون دراسة ضيقة وتكون النتائج قاصرة على مجموعة المرضى الذين يتصفون بهذه الصفات الخاصة ، والأفضل أن تشمل الدراسة مجالاً أوسع لكي تكون النتائج أكثر عمومية ، وذلك بتجريب نوعي المسكنات على مختلف المرضى بذلك المرض نأخذهم عمداً من الجنسين ومن مختلف المجموعات العمرية والظروف الصحية .

ومن الضروري إذا إيجاد حل وسط بين متطلب التجانس في وحدات التجريب ومتطلب تنويع هذه الوحدات ، وهما متطلبان متعارضان .

وقد وجد الحل بتصميم تجارب تعرف بتجارب القطاعات كاملة التعشية randomized complete blocks ( كلمة كاملة تعني أن كل مستوى من مستويات العامل يظهر نفس العدد من المرات في كل قطاع ) . وسنهتم هنا بحالة خاصة من هذه التجارب تعرف بالمقارنات التزاوجية وهي الحالة التي يكون فيها عامل التجريب ذا مستويين فقط ( $k = 2$ ) وسميت تزاوجية لأن كل وحدة من وحدات التجريب تتلقي أحد المستويين ، تتزوج مع وحدة مماثلة تتلقي المستوى الآخر . على أن الأسلوب الذي نتبعه في تناول وتحليل هذه الحالة يمتد إلى الحالات التي يكون فيها لعامل التجريب أكثر من مستويين .

والطريقة الإجرائية لهذه التجارب تعتمد على وضع وحدات التجريب مثني مثني في قطاعات يتألف كل منها من وحدتين تتلقي إحداهما المستوى الأول وتتلقي الأخرى المستوى الثاني ، بشرط أن تكون الودعتان في كل قطاع متشابهتين - بل قد تكون نفس الوحدة تعالج مرتين في وقتين مختلفين - بشرط أن يكون توزيع المستويين على الودعتين عشوائياً ( مثلاً بإلقاء قطعة من العملة ) في كل قطاع على حدة . أما الوحدات في القطاعات المختلفة فقد تكون متشابهة أو غير متشابهة .



إن هذا الإجراء يصون فعالية المقارنة داخل كل قطاع ويسمح في الوقت ذاته بتعدد الظروف بين مختلف القطاعات ، ويمكن النظر إلى الاستجابة في أى وحدة تجريبية على أنها مؤلفة من ثلاثة عناصر هي :

- ( أ ) تأثير مستوى المعالجة لعامل التجريب .
- ( ب ) تأثير الظروف داخل القطاعات ( ويعتبر عاملاً ثانياً ) .
- ( جـ ) مركبة عشوائية .

وتوضع استجابات وحدات التجريب كما يلي حيث س<sub>١</sub> ، س<sub>٢</sub> ، ... ترمزان إلى الاستجابات في القطاع للمستوى الأول والمستوى الثاني على الترتيب .

القطاع	المستوى الأول	المستوى الثاني
(١)	س <sub>١</sub>	س <sub>١</sub>
(٢)	س <sub>٢</sub>	س <sub>٢</sub>
(٣)	س <sub>٣</sub>	س <sub>٣</sub>
....	....	....
(ن)	س <sub>ن</sub>	س <sub>ن</sub>

بهذا التصور نكون بصدد حالة خاصة للتجارب ذوات العاملين غير المتفاعلين ، وبالتالي تسير عملية تحليل التباين كما في البند ( ٨ - ٦ - ١ ) السابق .

### مثال ( ٨ - ٨ ) :

أراد باحث أن يعرف ما إذا كان عرض الجزء الأسفل من وجه النبات أكبر في سن السادسة منه في سن الخامسة ، فاختار عينة عشوائية من ١٥ بتاً وقاس عرض الوجه وهن في الخامسة ثم أعاد القياس بعد عام ، وسجلت الأطوال بالاستيمتر كما يلي :

جدول (٨ - ١٠)

الأفراد (القطاعات)	(١) ٥ سنوات	(٢) ٦ سنوات	ك
( ١ )	٧,٣٣	٧,٥٣	١٤,٨٦
( ٢ )	٧,٤٩	٧,٧٠	١٥,١٩
( ٣ )	٧,٢٧	٧,٤٦	١٤,٧٣
( ٤ )	٧,٩٣	٨,٢١	١٦,١٤
( ٥ )	٧,٥٦	٧,٨١	١٥,٣٧
( ٦ )	٧,٨١	٨,٠١	١٥,٨٢
( ٧ )	٧,٤٦	٧,٧٢	١٥,١٨
( ٨ )	٦,٩٤	٧,١٣	١٤,٠٧
( ٩ )	٧,٤٩	٧,٦٨	١٥,١٧
(١٠)	٧,٤٤	٧,٦٦	١٥,١٠
(١١)	٧,٩٥	٨,١١	١٦,٠٦
(١٢)	٧,٤٧	٧,٦٦	١٥,١٣
(١٣)	٧,٠٤	٧,٢٠	١٤,٢٤
(١٤)	٧,١٠	٧,٢٥	١٤,٣٥
(١٥)	٧,٦٤	٧,٧٩	١٥,٤٣
ك معد محرر من س	١١١,٩٢	١١٤,٩٢	٢٢٦,٨٤ = ن ٣٠ ٧,٥٦ = س ١٧١٨,١٦٠٤ = م م س' س

لدينا عامل تحريب واحد هو العمر وهذا العامل له مستويان هما ٥ سنوات ، ٦ سنوات ونريد المقارنة بين متوسطي عرض الوجه في هذين المستويين . وتحسباً لوجود فروق بين أفراد العينة فإننا ندخل هؤلاء الأفراد كعامل ثان مع ملاحظة أن كل فرد يمثل

قطاعاً خضع لكل من مستويي العامل : وعلى أساس علم وجود تفاعل بين العاملين  
نسير في تحليل التباين كما في البند (٨ - ٦ - ١) لنجد ما يلي :

$$٢٢ \text{ (الكل)} = ١٧١٨,١٦٠٤ - \frac{٢٢٦٨٤}{٣٠} = ١٧١٥,٢١٢٨ - ١٧١٨,٦٠٤$$

$$٢,٩٤٧٦ = \text{حيث } \mathcal{P} = ٢٩$$

$$\text{م م (بين العمرين)} = \frac{٢١١١,٩٢ + ٢١١٤,٩٢}{١٥} - ١٧١٥,٢١٢٨$$

$$= ١٧١٥,٥١٢٨ - ١٧١٥,٢١٢٨$$

$$٠,٣٠٠٠ = \text{حيث } \mathcal{P} = ١$$

$$\text{م م (بين الأفراد)} = \frac{٢١٤,٨٦ + ٢١٥,١٩ + \dots + ٢١٥,٤٣}{٢} - ١٧١٥,٢١٢٨$$

$$= ٧١٧,٨٤٩٦ - ١٧١٥,٢١٢٨$$

$$٢,٦٣٦٨ = \text{حيث } \mathcal{P} = ١٤$$

$$\text{م م (الخطأ)} = ٢,٩٤٧٦ - (٢,٦٣٦٨ + ٠,٣٠٠٠)$$

$$= ٠,٠١٠٨ = \text{حيث } \mathcal{P} = ١٤$$

جدول (٨ - ١١)

مصدر التباين	م م	٢	قيمة التباين	ف
بين العمرين	٠,٣٠٠٠	١	٠,٣٠٠٠	٣٨٩,١١
بين الأفراد	٢,٦٣٦٨	١٤	٠,١٨٨٣	٢٤٢,٠٢
الخطأ	٠,٠١٠٨	١٤	٠,٠٠٠٧	
الكل	٢,٩٤٧٦	٢٩		

## الاستنتاج :

- (١) نسبة التباين للأعمار ذات دلالة عالية لأن  $F_{0.01} [١٤, ١] = ٠.٠١$  أصغر من ٩ ، مما يجعلنا نستنتج أن عرض وجه البنات في سن السادسة أوسع منه في سن الخامسة .
- (٢) كما أن الفروق بين أفراد العينة ذات دلالة عالية لأن  $F_{0.01} [١٤, ١٤] = ٠.٠١$  أصغر من ٤ .

## ملاحظة (١١) :

إذا كنا لم ندخل في اعتبارنا الاختلافات بين أفراد العينة وأجرينا عملية تحليل التباين على أساس أن لدينا تجربة ذات عامل واحد كما في البند (٨ - ٤) نحصل على جدول التباين الآتي :

جدول (٨ - ١٢)

مصدر التباين	م	مجموع التباين	ف
بين العمرين	١	٠,٣٠٠٠	٣,١٧
الخطأ	٢٨	٢,٦٤٧٦	٠,٠٩٤٦
الكلي	٢٩	٢,٩٤٧٦	

وهنا نجد أن نسبة التباين للأعمار ليست ذات دلالة مما يشير إلى عدم وجود فرق بين عرض الوجه في العمرين الخامسة والسادسة . وهذه النتيجة خاطئة وتختلف ما توصلنا إليه من قبل والسبب في ذلك عدم تجانس أفراد العينة ووجود فروق

يبين نشأ عنه اختلافات كبيرة كان يجب أن تستبعد ولكنها أضيفت إلى الاختلاف العشوائي فتضخم مجموع مربعات الخطأ وهذا بدوره دخل في مقام نسبة التباين فجعل هذه النسبة أصغر من أن تكون ذات دلالة ، وبالتالي اختفى الفرق بين مستويي عامل التجريب .

إن للمقارنات التزاوجية تطبيقات عديدة في مختلف ميادين البحث العلمي خاصة عند القياس أو الاختبار المتكرر لنفس المجموعة بعد فترة ما أو بعد حدث ما حيث يجري القياس « قبل وبعد » هذه الفترة أو هذا الحدث . ومن أمثلة ذلك اختبار قوة عضلات مجموعة من الأفراد ثم تعريضهم لتمرينات رياضية عنيفة ثم اختبار قوة عضلاتهم مرة أخرى . كذلك قياس خاصة ما لمجموعة من الكائنات الحية أو الأفراد ثم قياس هذه الخاصة لنفس المجموعة بعد مرحلة ما كما في المثال (٨ - ٨) الأخير .

كذلك تنتج المقارنات التزاوجية عند تقسيم وحدة ما إلى نصفين يتلقى أحدهما أحد مستويي عامل ما ويتلقى النصف الآخر المستوى الثاني الذي يمكن أن يكون مستوى المراقبة . وكمثال لذلك اختبار قوة نوعين من المضادات الحيوية بمقن أحدهما في الذراع الأيمن والآخر في الذراع الأيسر لنفس الشخص ثم قياس قطر البقع الحمراء الناتجة في كل من الحالتين .

ومن التصميمات التي تؤدي إلى مقارنات تزاوجية أيضاً إعطاء معالجتين إلى شخصين يشتركان في خبرة واحدة سواء كانت خبرة وراثية أو بيئية ، كإعطاء دواء إلى مجموعة من التوائم أو الأشقاء يتلقى أحدهما الدواء ولا يتلقاه الآخر ( مجموعة مراقبة ) .

### (٨ - ٧ - ١) اختبارات للمقارنات التزاوجية :

ذكرنا في الملاحظة (٤) بالبنء (٨ - ٤) أننا حين نتناول عاملاً ذا مستويين (ك = ٢) يمكن أن نستخدم اختباراً للمقارنة بين متوسطي هذين المستويين

وتكون النتيجة التي نحصل عليها من هذا الاختبار مطابقة تماماً للنتيجة التي نحصل عليها من تحليل التباين . غير أن الإحصاءة التي مرت بنا بالبند (٦ - ٦ - ٣) لا تصلح لهذا الغرض في حالة المقارنات التزاوجية لأن أحد شروط استخدام تلك الإحصاءة استقلال المجموعتين بينما نحن هنا بصدد مجموعتين مرتبطتين بل هما نفس المجموعة . أما الإحصاءة التي تصلح لذلك فتأخذ الصورة الآتية :

$$ت = \frac{\overline{ف} - (\overline{ف_1} - \overline{ف_2})}{\sqrt{\frac{ع}{ن}}} \quad \text{بدرجات حرية } \nu = ن - ١ \quad (١٣)$$

$$\text{حيث } \overline{ف_1} = \overline{س_1} - \overline{ص_1}$$

$$، \quad \overline{ف_2} = \overline{س_2} - \overline{ص_2} = \overline{س} - \overline{ص}$$

= متوسط الفروق بين استجابات المستويين

$$ع' = \frac{1}{ن - ١} = \overline{ف_1} - \overline{ف_2} = \text{تباين الفروق } \overline{ف_1}$$

$$(١٤) \quad \frac{1}{ن - ١} = [\overline{ف_1} - \overline{ف_2}] = \frac{1}{ن}$$

مع ملاحظة أن  $\sqrt{\frac{ع}{ن}}$  هو الخطأ المعياري لمتوسط الفروق مقدراً من العينة. والاختبار الذي نجره باستخدام الإحصاءة (١٣) يعطى نتيجة مطابقة تماماً لما تعطيه طريقة تحليل التباين إلا أنه لا يزودنا بمقياس لتباين القطاعات (الصفوف) . وفي المثال (٨ - ٨) الأخير يمكننا الحل كما يلي :



المجدول (٨ - ١٣)

الأفراد	ق - س - ص	ق	الأفراد	ق - س - ص	ق
١	٠,٢٠	٠,٠٤	٩	٠,١٩	٠,٠٣٦١
٢	٠,٢١	٠,٠٤٤١	١٠	٠,٢٢	٠,٠٤٨٤
٣	٠,١٩	٠,٠٣٦١	١١	٠,١٦	٠,٠٢٦٥
٤	٠,٢٨	٠,٠٧٨٤	١٢	٠,١٩	٠,٠٣٦١
٥	٠,٢٥	٠,٠٦٢٥	١٣	٠,١٦	٠,٠٢٥٦
٦	٠,٢٠	٠,٠٤	١٤	٠,١٥	٠,٠٢٢٥
٧	٠,٢٦	٠,٠٦٧٦	١٥	٠,١٥	٠,٠٢٢٥
٨	٠,١٩	٠,٠٣٦١			
				٣,٠٠	٠,٦٢١٦

$$\overline{ق} = \frac{٣}{١٥} = ٠,٢٠ \text{ من المستحتمر}$$

$$ع = \frac{١}{١٤} = (٠,٦٢١٦ - \frac{٣}{١٥}) = ٠,٠١٥٤٣$$

$$ع = ٠,٠٣٩٢٨١٠$$

$$ع = \sqrt[١٥]{٠,٠٣٩٢٨١٠} = ٠,٠١٠١٤٢٣$$

الفرض الصفري ف:  $\mu = \mu$

الفرض الآخر ف:  $\mu < \mu$

بحساب قيمة الإحصاءة (١٣) من بيانات العينة ( وعلى أساس صحة الفرض الصفري ) نجد أن

$$ت = \frac{0.20 - 0}{0.1014} = 19.724$$

بل درجات حرية عددها ١٤

وهذه القيمة ذات دلالة عالية مما يجعلنا نرفض الفرض الصفري ونستنتج أن عرض وجه البنات أوسع في سن السادسة منه في سن الخامسة .

### ملاحظات :

- (١) يمكن أن تثبت رياضياً أن مربع قيمة المتغيرات عند درجة الحرية ن يساوى قيمة المتغير ف عند درجتى الحرية ( ١ ، ن ) . ففى المثال الأخير نجد أن :  $ت^2 = 19.724 = 389.036$  . وهذا العدد يساوى العدد  $F = 389.11$  الذى ظهر بالجدول ( ٨ - ١١ ) فيما عدا الفرق الناشئ عن عمليات التقريب .
- (٢) فى مثل هذه الحال يكون لدينا متغيران غير مستقلين  $ص$  ،  $هـ$  ، إلا أننا فى استخدام الإحصاءة (١٣) نعتبر أن لدينا متغيراً واحداً ف حيث  $ف = س - ص$  . ولما كان المتغيران  $ص$  ،  $هـ$  هما متغيران مختلان متوسطهما  $م_1$  ،  $م_2$  وتباينهما  $\sigma_1^2$  ،  $\sigma_2^2$  على الترتيب ، فإن المتغير  $ف$  يكون متغيراً مختلاً متوسطه  $م_1 - م_2$  وهذا المتوسط يتعلم إذا كان الفرض الصفري صحيحاً . أما التباين  $\sigma^2$  فنقلره من العينة بالمقدار  $\sigma^2$  المعروف فى (١٤) بصرف النظر عن تساوى أو عدم تساوى التباينين  $\sigma_1^2$  ،  $\sigma_2^2$  . أى أننا فى استخدامنا للإحصاءة (١٣) لا نكون بحاجة لفرض تساوى هذين التباينين .

- (٣) إن حدى الثقة بدرجة ١ -  $\alpha$  للفرق  $م_1 - م_2$  بين متوسطى مجتمعى المتغيرين غير المستقلين  $ص$  ،  $هـ$  هما ( بالأسلوب المعتاد ) :

$$\bar{ق} \pm \frac{\sigma}{\sqrt{n}} ت_{\alpha/2, n-2}$$

ففى المثال الأخير نجد أن حدى الثقة بدرجة ٩٩٪ للفرق بين متوسط عرض وجه البنات فى سن السادسة ومتوسط عرضه فى سن الخامسة هما  $0.20 \pm 0.1014 \times 2.977$  أى  $0.170$  ،  $0.230$  من المستحتملات

## تمارين (٨ - ٤)

١ - أراد طبيب باحث أن يقرر ما إذا كان تعاطى قرص من مادة معينة يحدث تأثيراً جانبياً غير مرغوب فيه من حيث تخفيض ضغط الدم . وقد بدأت التجربة بقياس ضغط الدم لعينة عشوائية من ١٥ فرداً ثم إعطاء الأقراص لأفراد هذه العينة ، وانتهت بقياس ضغط الدم مرة أخرى بعد فترة معينة . سجلت القياسات كما يلى حيث س تعبر عن الضغط قبل تعاطى القرص ، ص تعبر عن الضغط لنفس الشخص بعد تعاطى القرص .

الفرد :	(١)	(٢)	(٣)	(٤)	(٥)	(٦)	(٧)	(٨)	(٩)	(١٠)	(١١)	(١٢)	(١٣)	(١٤)	(١٥)
س :	٧٠	٨٠	٧٢	٧٦	٧٦	٧٦	٧٢	٧٨	٨٢	٦٤	٧٤	٩٢	٧٤	٦٨	٨٤
ص :	٦٨	٧٢	٦٢	٧٠	٥٧	٦٦	٦٨	٥٢	٦٤	٧٢	٧٤	٦٠	٧٤	٧٢	٧٤

هل هذه البيانات تؤدى إلى الحكم بأن للأقراص تأثيراً فى تخفيض ضغط الدم ؟ استخدم كلا من طريقة تحليل التباين واختبارات وقارن بين التيجتين .

٢ - كان أحد الأطباء يشك فى أن الميزان الذى يزن به المرضى فى عيادته يعطى قراءات أعلى من القراءات التى يجدها المرضى عند استخدامهم للموازين التى فى منازلهم . ولاختبار ذلك طلب الطبيب من عشرة مرضى تسجيل أوزانهم بملايسهم الكاملة قبل مغادرتهم منازلهم إلى عيادته ثم قام بوزنهم فور وصولهم فحصل على البيانات الآتية حيث س تعبر عن الوزن فى العيادة ، ص تعبر عن الوزن بالمنزل

س :	١٥٣	١٤٦	٢٠٧	١٧٣	١٢٧	١١٢	١٨١	١٣٢	١٣٩	٢١٣
ص :	١٥٠	١٤٧	٢٠٣	١٧١	١٢٩	١١٠	١٧٩	١٣٢	١٣٨	٢١٠



وصفوف يمكن أن تمثل عناصرها مستويات عامل ثان ، ولدينا أيضاً حروف ا ، ب ، ج ، د ، ... يمكن أن تمثل مستويات عامل ثالث . ولذلك تصلح المربعات اللاتينية لدراسة التجارب ذوات الثلاثة العوامل بطريقة تعد امتداداً للطريقة التي استخدمناها بالبند ( ٨ - ٦ - ١ ) . وبالإضافة إلى الشروط المعتادة ، ينبغي توفر الشرطين الآتيين لاستخدام المربعات اللاتينية :

١ - أن يتساوى عدد مستويات كل من العوامل الثلاثة ، وبذلك يكون عدد الأعمدة = عدد الصفوف = عدد الحروف = ل ، ويكون عدد وحدات التجريب هو  $n = l^2$  .

٢ - ألا يكون هناك تفاعلات بين العوامل الثلاثة ، لأن هذا التصميم لا يقيس التفاعل نظراً لأن هناك مشاهدة واحدة فقط في كل خلية .

( يحسن ألا تستخدم المربعات اللاتينية التي تقل رتبها عن خمسة لعدم وجود معلومات كافية عن مدى حساسية المربعات اللاتينية لانحراف ظروف التجربة عن الفروض الموضوعية ) .

ولتحليل التباين من مربع لاتيني نوجد  $٢٢$  ( الكلى ) ،  $٢٢$  ( بين الأعمدة ) ،  $٢٢$  ( بين الصفوف ) كما في البند ( ٨ - ٦ - ١ ) وبالنسبة للعامل الثالث نوجد الاختلاف الناشئ عنه من الصيغة :

$$٢٢ \text{ (الحروف)} = \frac{\sum_{j=1}^l \sum_{i=1}^l x_{ij}^2}{l} - \frac{\sum_{i=1}^l x_i^2}{n}$$

حيث  $\sum$  هو مجموع القيم المرتبطة بالحرف  $s$  (  $s = ١, ٢, ٣, \dots$  ) . أما  $\sum$  ( الخطأ ) فيحسب بطرح مجموع الاختلافات الناشئة عن العوامل الثلاثة من الاختلاف الكلى .

### مثال ( ٨ - ٩ ) :

لاختبار تأثير ٣ عوامل على المحصول وهى تأثير اختلاف الأسمدة ( ٥ أنواع ) وتأثير اختلاف التربة في اتجاهين متعامدين ، قسمت قطعة أرض إلى ٢٥ حوضاً متساوية المساحة ومرتبطة في ٥ صفوف موازية لأحد الاتجاهين ، ٥ أعمدة موازية للاتجاه المتعامد عليه ، ثم وزعت الأنواع الخمسة من الأسمدة عشوائياً على الأحواض بحسب خطة مربع لاتيني من الرتبة الخامسة ، ونتج المربع اللاتيني الآتي حيث تشير الحروف إلى الأسمدة وتشير الصفوف والأعمدة إلى الاتجاهين المتعامدين وتشير الأعداد إلى الكميات الناتجة من المحصول ( مطروح ١٠ من كل منها ) والمطلوب بحث تأثير كل من هذه العوامل الثلاثة .

١,٦-	أ	١,١-	د	٤,٢-	هـ	٢,٠	ب	٤,٣	ح	١,٦-
٧,٥-	ح	٣,٥-	ب	١,٣-	أ	٢,٤-	هـ	٢,١-	د	٧,٥-
٢,٦	د	٧,٩	ح	١,٠-	ب	١,٥-	أ	٢,٩-	هـ	٢,٦
٣,١	هـ	١,٢-	أ	٥,٧	ح	١,١	د	٢,٦-	ب	٣,١
١٣,٤	ب	١,٢-	هـ	٤,٣	د	٨,٤	ح	٠,١	أ	١٣,٤
١٠,١-	ب	٢,٢	٣,٥	٧,٦	٣,٢-	١٠,١-				

الحل :

بالنسبة للأسمدة ( الحروف ) نجد أن المجاميع  $\sum$  كما على :

٢٢ : أ = -٢,٥ ، ب = -٣,٢ ، ج = ١,٢٨ ، د = ٣,٢ ، هـ = -٩,١١

$$\frac{1}{2} - \frac{1}{2} = \text{مجموع مساحات} \quad (الكل) \quad ٢٢$$

$$٢٤ = ٢٨١,١٨ = \frac{١٠٠}{٢٥} - ٢٨٥,١٨ =$$

$$\frac{1}{2} - \frac{1}{2} = \text{مجموع المساحات} \quad (بين الأعمدة) \quad ٢٢$$

$$٤ = ١٣,٠٢ = ٤ - ١٧,٠٢ =$$

$$\frac{1}{2} - \frac{1}{2} = \text{مجموع المساحات} \quad (بين الصفوف) \quad ٢٢$$

$$٤ = ٤٦,٩٥ = ٤ - ٥٠,٩٥ =$$

$$\frac{1}{2} - \frac{1}{2} = \text{مجموع المساحات} \quad (بين الأعمدة) \quad ٢٢$$

$$٤ = ١٩٠,٨٩ = ٤ - ١٩٤,٨٩ =$$

$$(١٩٠,٨٩ + ٤٦,٩٥ + ١٣,٠٢) - ٢٨١,١٨ = \quad (الخطأ) \quad ٢٢$$

$$١٢ = ١٢ - ٢٤ = \text{حيث} \quad ٣٠,٣٢ =$$

$$٥,٤١ = [١٢, ٤]٠,٠١ \quad \text{ف} \quad ٣,٣٦ = [١٢, ٤]٠,٠٥$$

الامتحان : انظر الجدول (٨ - ١٤) .

(١) الاختلاف الناشئ عن الأعمدة ذو دلالة عالية .

(٢) الاختلاف الناشئ عن الأعمدة ليس له دلالة عند المستوى ٠,٠٥ .

(٣) الاختلاف الناشئ عن الصفوف ذو دلالة عند المستوى ٠,٠٥ وليس ذو

دلالة عند المستوى ٠,٠١

جداول (٨ - ١٤)

مصدر البيانات	مجموع للربح	درجات الحرية	تقدير التباين	د
بين الأعمدة	١٣,٠٢	$1 - 1 = 0$	٢,٢٥٥	١,٣
بين الصفوف	٤٦,٩٥	$1 - 1 = 0$	١١,٧٣٧	٤,٦
بين الأعمدة الخطأ	١٩٠,٨٩	$1 - 1 = 0$	٤٧,٧٢٢	١٨,٩
	٣٠,٣٢	$(2-1)(1-1) = 0$	٢,٥٢٧	
المجموع	٢٨١,١٨	$24 - 1 = 23$		

## تأريين (٨ - ٥)

(١) في دراسة في أبحاث السوق كان المطلوب اختبار تأثير ثلاثة عوامل على مبيعات نوع معين من الغذاء في قطر ما وهذه العوامل هي :

(١) طرق التغليف - ٤ مستويات : أ ، ب ، ج ، د

(٢) المناطق المختلفة في القطر - ٤ مستويات .

(٣) طرق التشجيع على الشراء - ٤ مستويات : نسبة تخفيض ، يانصيب ، كوبونات ، هدايا .

وقد أجريت تجربة باستخدام تصميم مربع لاتيني من الرتبة الرابعة وسجلت المبيعات في أسبوع بعشرات الآلاف من الدولار كالآتي :



المناطق	تخصيص	ياتصیب	كروانات	هدايا	المجموع
(١)	أ ٤٨	ب ٣٨	ج ٤٢	د ٥٣	١٨١
(٢)	ب ٣٩	ج ٤٣	د ٥٠	أ ٥٤	١٨٦
(٣)	ج ٤٢	د ٥٠	أ ٤٧	ب ٤٤	١٨٣
(٤)	د ٤٦	أ ٤٨	ب ٤٦	ج ٥٢	١٩٢
المجموع	١٧٥	١٧٩	١٨٥	٢٠٣	٧٤٢

$$ع = ١٩٧ \text{ ب} = ١٦٧ \text{ ج} = ١٧٩ \text{ د} = ١٩٩$$

ابحث دلالة تأثير كل من العوامل الثلاثة .

(٢) في المثال (٨ - ٩) أثبت أن الخطأ المعياري

$$ع = \frac{1}{n} + \frac{1}{n} \sqrt{\frac{2}{n}} \text{ للفرق بين متوسطين هو } ١,٠٠٥$$

وعن ثم بين أن الفرق الذي يقل عن ٢,٢ بين أي متوسطين لا يكون ذا دلالة عند المستوى ٠,٠٥ ومن ثم برهن أن :

(أ) متوسط المحصول الناتج من السماد ج أكبر جوهرياً من ذلك الناتج من أي نوع آخر من السماد .

(ب) متوسط المحصول الناتج من السماد د أكبر جوهرياً من ذلك الناتج من السماد هـ .

## (٨ - ٩) معالجة الانحرافات عن افتراضات التحليل :

إن سلامة ما نجره من تحليل وما نخرج به من نتائج تتوقف على توفر الافتراضات المذكورة في البند (٨ - ٤ - ١) . ولكن ماذا يكون موقفنا إذا لم تكن بعض هذه الافتراضات مستوفاة ؟ هذا ما سناقشه كما يلي :

### (أ) الفراض الاعتدالية :

إن افتراض اعتدال المجتمعات التي نتناولها في تحليل التباين هو افتراض رئيسي . ولكن نظرا لأننا في هذا التحليل ( بالتموذج ثابت التأثيرات ) نبحت في الفروق بين المتوسطات فإن هذا الافتراض يمكن أن تحلل منه دون خطورة بشرط أن تكون العينات كبيرة كبرا كافيا ( أى لا يقل حجم كل منها عن ٣٠ ) . وذلك لأنه حسب نظرية النهاية المركزية - انظر ملاحظة البند (٦ - ٣) - إذا كانت الاستنتاجات المتعلقة بالأوساط الحسابية صحيحة حين تكون المجتمعات معتدلة فإنها تظل صحيحة حين تكون المجتمعات غير معتدلة بشرط أن تكون العينات كبيرة كبرا كافيا . وكلما اشتد انحراف المجتمع عن الاعتدالية كلما وجب علينا زيادة حجم العينات .

### (ب) الفراض تساوى التباينات :

يتضمن الافتراض الثاني من افتراضات تحليل التباين أن تكون مجتمعات أقسام أو مستويات عامل التجريب متساوية التباين . يمكن التجاوز عن هذا الافتراض بشرط أن تكون العينات متساوية الحجم لأنه في هذه الحال لا يكون التحليل حساسا للانحرافات الصغيرة عن افتراض تساوى التباينات . أما إذا لم تكن العينات متساوية الحجم فإن اختلاف تباينات مجتمعاتها يكون ذا عواقب وخيمة على صحة الاستنتاج . ولهذا يفضل سحب عينات متساوية الحجم كلما أمكن ذلك .

### (ج) افترض استقلال الأخطاء :

يتطلب الافتراض الثالث أن تكون أخطاء التجريب مستقلة عن بعضها وعن مستويات المعالجة ، وهذا أمر بالغ الأهمية لسلامة استخدام اختبار ف في تحليل التباين . وعدم توفر هذا الشرط يمكن أن يؤدي إلى خطأ جسيم في الاستنتاج . ولذلك وجب اتخاذ القدر الكافي من الحيلة في عملية التجريب بحيث نضمن استقلال المشاهدات داخل وبين المجموعات ، ويساعدنا على ذلك تطبيق مبدأ العشوائية في كافة جوانب التجربة وقد سبق الإشارة إلى ذلك .

وفي كثير من الأحيان يمكن تصحيح بعض الانحرافات عن الفروض الموضوعية باستخدام أحد التحويلات المناسبة كالمذكورة في البند (٤ - ٥) . على أنه إذا فشلنا في توفير هذه الفروض فلا مفر من اللجوء في التحليل إلى أحد الطرق غير البارامترية التي سنتناولها في الفصل الرابع عشر .

### (٨ - ١٠) عودة إلى مقارنة المتوسطات :

نعلم أن تحليل التباين ما هو إلا الخطوة الأولى للدراسة نتائج التجربة وأن الخطوة الثانية ( في النموذج ثابت التأثيرات ) هي مقارنة متوسطات أقسام المعالجة واختبار ما قد يكون بينها من فروق . ولقد قدمنا في البند (٨ - ٥) طريقة لاختبار هذه المقارنات في حالتها المقارنات القبلية والبعدي . ونقدم الآن أسلوباً أو مدخلاً آخر يسفر عن نفس الصيغ والاختبارات السابق تقديمها ولكن في صور مختلفة وبدلالة متغيرات أخرى ، وبالتالي يسفر عن نفس الاستنتاجات . على أن دراسة هذا المدخل تعمق مفهوم المقارنات وتضفي عليها معاني مفيدة . وفي هذا البند والبند الثلاث التالية تقدم بعض التعاريف الأساسية في هذا المدخل ونبدأ بالتعريف الآتي :

## تعريف (١) : المقارنة (أو المتضادة)

### COMPARISON (or CONTRAST)

#### (أ) المقارنة بين متوسطات مجتمعات :

لتكن  $\mu_1, \mu_2, \dots, \mu_k$  متوسطات  $k$  من المجتمعات . ولتكن  $\alpha_1, \alpha_2, \dots, \alpha_k$  أعدادا ثابتة ليست جميعها أصفارا . إن أى تعبير خطى  $\psi$  فى هذه المتوسطات :

$$\psi = \alpha_1 \mu_1 + \alpha_2 \mu_2 + \dots + \alpha_k \mu_k = \text{مجموع } \alpha_i \mu_i \quad (١٦)$$

يسمى مقارنة (خطية) بين متوسطات هذه المجتمعات إذا توفر الشرط الآتى :

$$\alpha_1 + \alpha_2 + \dots + \alpha_k = 0 \text{ أى } \sum \alpha_i = 0$$

ويتضمن هذا الشرط أن تكون بعض المعاملات  $\alpha_i$  (وتسمى أوزانا) موجبة والبعض الآخر سالبة . والمعتاد اختيار هذه الأوزان فى أى مقارنة بحيث يكون مجموع الأوزان الموجبة مساويا للواحد وبالتالي يكون مجموع الأوزان السالبة مساويا للعدد -١ . وهذا الإجراء ممكن دائما وعيه أنه يجعل الأوزان فى صور كسرية فى أغلب المقارنات ولذلك لا يفضلها بعض الباحثين . ومن أمثلة المقارنات الخطية ما يلى :

$$\mu_1 - \mu_2, \mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4, \mu_1 - \mu_2 - \mu_3 - \mu_4$$

#### (ب) المقارنة بين متوسطات عينات :

لتكن  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$  متوسطات  $k$  من العينات المأخوذة من

المجموعات موضع الدراسة . ولتكن  $a_1, a_2, \dots, a_n$  أعدادا ثابتة ليست جميعها أصفارا . إن أى تعبير خطى  $\hat{y}$  فى هذه المتوسطات :

$$\hat{y} = a_1 \bar{y}_1 + a_2 \bar{y}_2 + \dots + a_n \bar{y}_n \quad (17)$$

$$(18) \quad \text{حيث } \bar{y} = 0$$

يسمى مقارنة بين متوسطات هذه العينات .

ويمكن إثبات أن توفر الشرط (١٨) يجعل قيمة أى مقارنة بين متوسطات العينات مستقلة عن قيمة المتوسط العام  $\bar{y}$  لهذه العينات . وهذا أمر هام لأننا فى مقارنة متوسطات المجموعات ينبغى ألا يكون لمتوسطات العينات أى علاقة بقيمة  $\bar{y}$  التى نستخدمها لتقدير المتوسط العام  $\bar{y}$  للمجتمع .

### (ج) المقارنة بين مجاميع عينات

بالمثل ، لتكن  $a_1, a_2, \dots, a_m$  مجاميع  $k$  من العينات ولتكن  $a_1, a_2, \dots, a_m$  أعدادا ثابتة ليس جميعها أصفارا . إن أى تعبير خطى  $\hat{y}$  فى هذه المجاميع :

$$\hat{y} = a_1 \bar{y}_1 + a_2 \bar{y}_2 + \dots + a_m \bar{y}_m \quad (19)$$

$$\text{حيث } \bar{y} = 0$$

يسمى مقارنة بين مجاميع هذه العينات .

وفى تحليل التباين ، المعتاد استخدام المقارنات بين المجاميع وليس بين المتوسطات لأن ذلك يخفف من بعض عمليات القسمة والتقريب . غير أننا سنتناول هنا المقارنات بين المتوسطات لأننا أساسا ندرس هذه المتوسطات وبالتالي فإن تناولها يكون أكثر قدرة

على ما نريد توضيحه من مفاهيم . وعلى أية حال فإن ما سنقدمه من صيغ تخص المقارنات بين المتوسطات يمكن تحويلها إلى صيغ تخص المقارنات بين المجاميع بمجرد ضرب كل معامل  $\lambda$  في الحجم  $N$  للعينة الخاصة به .

( يلاحظ أن رمز المقارنة  $\lambda/\lambda$  أو  $\lambda/\lambda^*$  هو عبارة عن عدد واحد لأنه يتركب من مجموع أعداد ) .

إن أى تساؤل عن بعض أو كل متوسطات أقسام المعالجة يمكن وضعه في صورة مقارنة بين هذه المتوسطات ولا يستلزم الأمر إلا وضع وزن مناسب لكل متوسط بحيث يتحقق الشرط (١٨) ، مع ملاحظة إعطاء الوزن صفر للمتوسطات التي لا تدخل في المقارنة .

مثال (٨ - ١٠) :

في المثال (٨ - ١) كان لدينا خمسة أقسام حجم كل منها ١٠ ومتوسطاتها كالآتي :

(١) جلوكوز (٢) فركتوز (٣) جلوكوز + فركتوز (٤) سكروز (٥) مراقبة

٥٩,٣      ٥٨,٢      ٥٨      ٦٤,١      ٧٠,١

وفي الأمثلة (٨ - ٢) و (٨ - ٣) و (٨ - ٤) التابعة لهذا المثال كان المطلوب الإجابة عن التساؤلات القبلية الآتية :

(١) هل متوسط مجموعة السكريات الأربعة مجتمعة يختلف عن متوسط مجموعة المراقبة ؟

(٢) هل متوسط مجموعة السكريات النقية (١) و (٢) و (٤) مجتمعة يختلف عن متوسط السكر الخليط ؟

(٣) هل متوسطات مجموعات السكريات (١) و(٢) و(٣) و(٤) واحدة ؟

إن هذه التساؤلات يمكن صياغتها على هيئة مقارنات بين المتوسطات كالآتي ، مع ملاحظة أنه يمكن إعطاء أوزاناً أخرى تتناسب مع الأوزان المأخوذة هنا ومع ملاحظة ضرورة توفر الشرط (١٨) :

$$\frac{1}{4} = \frac{1}{4} \mu - \frac{1}{4} \mu + \frac{1}{4} \mu - \frac{1}{4} \mu + \frac{1}{4} \mu - \frac{1}{4} \mu$$

$$\frac{1}{4} = \frac{1}{4} \mu - \frac{1}{4} \mu + \frac{1}{4} \mu - \frac{1}{4} \mu + \frac{1}{4} \mu - \frac{1}{4} \mu \times \text{صفر}$$

أما التساؤل الثالث فيتضمن مقارنتين يمكن كتابتهما بصور مختلفة منها :

$$\frac{1}{4} \mu - \frac{1}{4} \mu = \frac{1}{4} \mu$$

$$\frac{1}{4} \mu - (\frac{1}{4} \mu + \frac{1}{4} \mu) = \frac{1}{4} \mu$$

وعادة ما تكتب المقارنات المناظرة بين متوسطات العينات في جدول كالآتي :

جدول (٨ - ١٥)

المقارنات المتعلقة بالأطعمة (٨ - ٧) و(٨ - ٣) و(٨ - ٤)

المقارنة	المتوسط				
	(١)	(٢)	(٣)	(٤)	(٥)
	٥٩,٣	٥٨,٢	٥٨	٦٤,١	٧٠,١
١- : سكريات ضد مرقة	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	١-
٢- : سكريات نقية ضد سكر خليط	$\frac{1}{4}$	$\frac{1}{4}$	١-	$\frac{1}{4}$	٠
٣- : جلوكوز ضد فركتوز	١	١-	٠	٠	٠
٤- : (جلوكوز+فركتوز) ضد سكروز	$\frac{1}{4}$	$\frac{1}{4}$	٠	١-	٠

من هذا الجدول يسهل إيجاد قيم المقارنات الأربع كالآتي :

$$\begin{aligned} \hat{p}_1 &= \frac{1}{4} (59,3 + 58,2 + 58 + 64,1) - 70,1 = 10,2 \\ \hat{p}_2 &= \frac{1}{4} (59,3 + 58,2 + 64,1 + 58) - \text{صفر} = 2,0222 \\ \hat{p}_3 &= 59,3 - 58,2 + \dots + \dots = 1,1 \\ \hat{p}_4 &= \frac{1}{4} (59,3 + 58,2 + \dots + 64,1) - \dots = 0,35 \end{aligned}$$

#### (٨ - ١١) المقارنات القبلية :

كما سبق القول ، يفضل اختيار المقارنات القبلية بحيث تكون مستقلة إحصائيا عن بعضها ، وذلك لكي تكون المعلومات الناتجة من أى مقارنة ذات قيمة تخصها وحدها ولا تتداخل مع المعلومات الناتجة من أى مقارنة أخرى . ولذلك يهمننا أن نعرف قاعدة نستدل بها على استقلال أو عدم استقلال المقارنات .

#### (٨ - ١١ - ١) معيار استقلال مقارنتين :

يمكن البرهنة رياضيا على النظرية الآتية :

لتكن  $p_1, p_2, \dots, p_k$  هي  $k$  من المتغيرات العشوائية المستقلة التي لها توزيعات معدلة وتباين مشترك . ولتكن  $q_1, q_2$  هما المقارنتان :

$$\begin{aligned} q_1 &= p_1 - p_2 + p_3 - p_4 + \dots + p_k \\ q_2 &= p_1 + p_2 - p_3 + p_4 - \dots + p_k \end{aligned}$$

يكون المتغيران العشوائيان  $q_1, q_2$  مستقلين إحصائيا إذا توفر الشرط الآتي :

$$\begin{aligned} p_1 + p_2 + \dots + p_k &= 0 \\ \text{أى} \quad \sum_{i=1}^k p_i &= 0 \end{aligned} \quad (20)$$

(لاحظ أن هذا الشرط يعتمد فقط على الأوزان ولا يعتمد على المتغيرات) .



إن هذه النظرية تمنحنا قاعدة أو معيارا للكشف عن استقلال المقارنات . فإذا كان لدينا  $k$  من المجتمعات المعتدلة التي تشترك في التباين وسحبنا منها  $k$  من العينات المستقلة التي لها نفس الحجم فإنه تطبيقا لهذه النظرية تكون أى مقارنتين  $\hat{\mu}_i$  و  $\hat{\mu}_j$  بين متوسطات العينات مستقلتين إذا توفر الشرط (٢٠) .

ويقال لمقارنتين ينطبق عليهما معيار الاستقلال إنهما متعامدتان orthogonal . وتعامد مقارنتين هنا يعنى استقلالهما إحصائيا ، ولذلك تستخدم كلمتا « التعامد » و « الاستقلال » كمتبادلتين ما دمنا نتعامل مع متغيرات مستقلة وتوزيعات معتدلة تشترك في التباين .

وحيث تكون العينات مختلفة الأحجام ،  $n_i$  ترمز إلى حجم العينة  $i$  فإن معيار الاستقلال يتخذ الصيغة الآتية :

$$S = \frac{1}{n_1} + \frac{1}{n_2} + \dots + \frac{1}{n_k}$$

$$S = \frac{1}{n} \quad \text{أى} \quad (21)$$

مثال (٨ - ١١) :

اختبر استقلال المقارنات المدونة بالجدول (٨ - ١٥) ، على فرض توفر شروط استقلال المتوسطات واعتدال المجتمعات وتساوى تبايناتها .

الحل :

بالجدول أربع مقارنات وإذن هناك  $n_{ij} = 6$  أزواج من المقارنات يراد اختبار استقلالها . ومع ملاحظة تساوى أحجام العينات نجد من الصيغة (٢٠) ما يلى :

$$S = \frac{1}{6} \times (1) + \frac{1}{6} \times (1) + \frac{1}{6} \times (1) + \frac{1}{6} \times (1) : \hat{\mu}_1 \text{ و } \hat{\mu}_2 \text{ للمقارنتين}$$

وإذن هاتان المقارنتان مستقلتان .

$$\text{للمقارنتين } \hat{\sigma}_1^2 \text{ و } \hat{\sigma}_2^2 : \frac{1}{\sigma_1^2} \times 1 + \frac{1}{\sigma_2^2} \times (-1) = 0$$

وإذن هاتان المقارنتان مستقلتان .

بالمثل نجد أن الأزواج الأربعة الباقية مستقلة ، وبذلك تكون المقارنات الستة مستقلة 'مثنى مثنى' . تحقق من ذلك .

(٨ - ١١ - ٢) اختبار المقارنات القبلية :

يستخدم اختبار المقارنات القبلية على الحقيقتين الآتيتين اللتين يمكن إثباتهما رياضياً .

(أولاً) المقارنة  $\hat{\sigma}_1^2$  بين متوسطات العينات هي تقدير غير متحيز للمقارنة  $\sigma^2$  بين متوسطات المجتمعات التي أخذت منها العينات والتي تحمل نفس الأوزان .  
(ثانياً) على فرض استقلال المتوسطات ، وإذا كان  $\sigma^2$  هو التباين المشترك للمجموعات فإن الخطأ المعياري للمقارنة  $\hat{\sigma}_1^2$  بين المتوسطات هو

$$\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (٢٢)$$

حيث  $n$  حجم العينة .

ونظراً لأن التباين  $\sigma^2$  يكون عادة غير معروف فإننا نقدره من البيانات المشاهدة بنفس طريقة التقدير في تحليل التباين أى بواسطة  $\hat{\sigma}_1^2$  وهو تباين خطأ التجريب . وبذلك يكون تقدير الخطأ المعياري للمقارنة  $\hat{\sigma}_1^2$  بين المتوسطات هو  $\hat{\sigma}_1^2$  حيث :

$$\hat{\sigma}_1^2 = \hat{\sigma}_1^2 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (٢٣)$$

من هاتين الحقيقتين وحين تتوفر شروط اعتدالية المجموعات وتساوى تبايناتها واستقلال المتوسطات - وهذا ما نفترضه عادة في تحليل التباين - فإنه حسب البند (٦ - ٦) يكون للإحصاءة

$$(٢٤) \quad \frac{\hat{\psi} - \psi}{\sqrt{E}} = T$$

توزيع ت بدرجات حرية عددها هو عدد درجات حرية  $E$  والذي نرمز له بالرمز  $\nu$ .

كما أنه من البند (٦ - ٦ - ٢) تكون الفترة

$$(٢٥) \quad (\hat{\psi} - \sqrt{E} \cdot T_{\alpha/2}, \hat{\psi} + \sqrt{E} \cdot T_{\alpha/2})$$

هي فترة ثقة بدرجة  $(1 - \alpha)$  للمقارنة  $\psi$ .

كل من اختبار ت بالصورة (٢٤) والفترة (٢٥) يصلح لاختبار أى فرض عن قيمة المقارنة  $\psi$ . وبالرغم من أن اختبار ت هو أصلاً ، كما نعلم ، اختبار للفرق بين متوسطى عينتين مستقلتين أى للمقارنات التى على الصورة  $\hat{\psi} = \bar{y}_1 - \bar{y}_2$  إلا أنه يصلح هنا أيضاً لاختبار أى مقارنة مهما كان عدد المتوسطات الداخلة فيها . وهذا استثناء فى استخدام اختبارات السابق دراسته . وبالنسبة للفترة (٢٥) ، إذا افترضنا قيمة معينة للمقارنة  $\psi$  ولم تقع هذه القيمة فى هذه الفترة فإننا ، كالعتاد ، نرفض الفرض  $\psi = 0$  عند مستوى الدلالة  $\alpha$  ( اختبار ذو جانبيين ) .

وفى معظم الحالات يكون المطلوب اختبار الفرض الصفرى  $\psi = 0$  ضد الفرض  $\psi \neq 0$  . وفى هذه الحالة تأخذ الإحصاءة (٢٤) الصيغة الآتية :

$$(٢٦) \quad \frac{\hat{\psi}_E}{\psi_E} = \text{ت} \quad \text{بلرجة حرية } \psi_E$$

$$(٢٧) \quad \frac{\hat{\psi}_E}{\psi_E} = \text{أو الصيغة المكافئة ف} \quad \text{بلرجتى حرية } \psi_E$$

وفى هذه الحالة أيضا نرفض الفرض الصفري  $\psi = 0$  عند مستوى الدلالة  $\alpha$  إذا كانت الفترة (٢٥) لا تحتوى الصفر .

مثال (٨ - ١٢) :

أجب عن التساؤل الأول المطروح بالمثال (٨ - ١٠) مستخدما مستوى الدلالة ٠,٠١ مع تذكر أننا وجدنا فى تحليل التباين أن تباين خطأ التجريب هو  $\psi_E = ٥,٤٦$  بلرجات حرية  $\psi_E = ٤٥$  وأن كلا من المتوسطات بنى على ١٠ مشاهدات

الحل :

التساؤل المطلوب يشير إلى المقارنة  $\psi$  بين متوسط أقسام السكريات مجمعة ومتوسط قسم المراقبة .

الفرض الصفري :  $\psi = 0$  والفرض الآخر  $\psi \neq 0$  .  
وجدنا أن  $\hat{\psi} = ١٠,٢$

من (٢٣) :  $\psi_E = \psi_E \neq \frac{\hat{\psi}_E}{\psi_E}$  حيث  $\psi = ٤٥$

$$[ \left( \frac{1}{1} \right) + \left( \frac{1}{4} \right) + \left( \frac{1}{4} \right) + \left( \frac{1}{4} \right) ] \frac{1}{1} \times ٥,٤٦ = ٠,٦٨٢٥ =$$

$$\text{من (٢٧) : في } = \frac{(10,2-)}{0,6825} = 102,448$$

وبما أن  $F_{[40, 1], 0,01} = 7,31$  نرفض الفرض الصفري عند المستوى ٠,٠١ .  
 يلاحظ أن قيمة  $F$  الناتجة هي نفس قيمة  $F$  التي سبق أن توصلنا إليها في المثال (٨ - ٢) .

ويمكن بطبيعة الحال استخدام اختبارات بالصيغة (٢٦) حيث نجد أن القيمة التي تنتج تساوى الجذر التربيعي لقيمة  $F$  التي حصلنا عليها . كما يمكن استخدام فترة الثقة (٢٥) كالآتي :

$$\text{الحد الأدنى للفترة} = 10,2- - 2,63 \times 0,6825\sqrt{7} = 12,37 -$$

$$\text{الحد الأعلى للفترة} = 10,2- + 2,63 \times 0,6825\sqrt{7} = 8,03 -$$

إذن الفترة (  $12,37 -$  ،  $8,03 -$  ) هي فترة ثقة بدرجة ٩٩٪ للمقارنة  $\frac{1}{\sqrt{2}}$  .  
 وبما أنها لا تحوى الصفري نرفض الفرض  $\frac{1}{\sqrt{2}}$  = ٠ عند المستوى ٠,٠١ .

إن الصيغ (٢٤) و (٢٥) و (٢٦) و (٢٧) لاختبار المقارنات القبلية التي وضعت أصلا للتجارب ذوات العامل الواحد تصلح كذلك للتجارب ذوات العاملين أو أكثر طالما كان النموذج المستخدم هو النموذج ثابت التأثيرات .

مثال (٨ - ١٣) :

أجب عن التساؤلات القبلية الثلاثة المطروحة بالمثال (٨ - ٧) علما بأن  $\chi^2_{ع} = 3,13$  بدرجات حرية ٤٥ وأن كلا من المتوسطات بنى على ٨ مشاهدات .

الحل :

نضع هذه التساؤلات على هيئة مقارنات كما بالجدول (٨ - ١٦) الآتي :

جدول (٨-١٦)

المقارنات المسقة بالمثال (٨-٧)

المقارنة				المتوسط
(٤)	(٣)	(٢)	(١)	
٢,٢٥	٣,٨٧٥	٥,٥	١٢,٦٢٥	
١-	.	.	١	٢٨ درجة الحرارة ضد ٨ درجة الحرارة $\uparrow$
.	١-	١	.	٢١ درجة الحرارة ضد ١٤ درجة الحرارة $\uparrow$
$\frac{1}{4}$	$\frac{1}{4}$ -	$\frac{1}{4}$ -	$\frac{1}{4}$	الدرجتان (٢٨ و ٢١) ضد الدرجتان (١٤ و ٨) $\uparrow$

$$١٠,٣٧٥ = ٢,٢٥ \times ١ - ١٢,٦٢٥ \times ١ = \uparrow$$

$$٠,٧٨٢٥ = (١+١) \frac{1}{8} \times ٣,١٣ = \uparrow$$

$$٦٣٧,٥٥٩ = ٠,٧٨٢٥ / (١٠,٣٧٥) = \text{في}$$

بما أن  $٧,٣١ = [٤٥٠,١]٠,٠١$  فإن  $\text{في}$  تكون ذات دلالة عالية

ونرفض الفرض الصفري أن  $\uparrow = ٠$

بالمثل نجد أن

$$\uparrow = ١,٦٢٥ \text{ و } \uparrow = ٠,٧٨٢٥ \text{ و } \text{في} = ٣,٣٧٥$$

وإذن نقبل الفرض أن  $\uparrow = ٠$

كذلك

$$\uparrow = ٢,٧٥ \text{ و } \uparrow = ٠,٣٩١٢٥ \text{ و } \text{في} = ٦٩,٣٢٩$$

وإذن نرفض الفرض أن  $\uparrow = ٠$

لاحظ أن هذه هي نفس النتائج التي توصلنا إليها في المثال (٨-٧).

# (٨ - ١٢) تجزئة مجموع المربعات بين أقسام المعالجة :

إن تحليل التباين بالمودج ثابت التأثيرات يكافئ فصل البيانات إلى مجموعة من المقارنات المستقلة ، إذ أن كل درجة من درجات الحرية تصاحب معالجة ما يناظرها مقارنة ما بين المتوسطات . ولتوضيح ذلك نبدأ بالتعريف الآتي :

تعريف (٢) : مجموع مربعات مقارنة :

إذا كانت  $\hat{\psi} = 1, \bar{s}_1 + 1, \bar{s}_2 + \dots + 1, \bar{s}_p$  مقارنة بين متوسطات العينات فإن مجموع مربعات هذه المقارنة يعرف كالآتي :

$$\hat{\psi}^2 = \frac{1}{n} \sum_{i=1}^p (\bar{s}_i - \bar{s})^2 \quad (٢٨)$$

ويمكن إثبات أن لهذا المجموع درجة واحدة من درجات الحرية .

فمثلا ، في المثال (٨ - ٢) الذي يتناول المقارنة  $\hat{\psi}$  بين أقسام السكريات مجتمعة ضد قسم المراقبة نجد ما يلي :

$$\hat{\psi}^2 = 10,2 - 10,2 = 0$$

$$\hat{\psi}^2 = \frac{1}{10} \left[ 1^2 + \left(-\frac{1}{4}\right)^2 + \left(-\frac{1}{4}\right)^2 + \left(-\frac{1}{4}\right)^2 + \left(-\frac{1}{4}\right)^2 \right] = 0,125$$

$$\therefore \hat{\psi}^2 = (10,2 - 10,2) / 0,125 = 832,32 \text{ بدرجة حرية واحدة .}$$

وهذا العدد بالضبط هو الذي وجدناه في المثال (٨ - ٢) .

كذلك ، في المثال (٨ - ٣) الذي يتناول المقارنة  $\hat{\psi}$  بين السكريات النقية والسكر الخليط نجد ما يلي :

$$\hat{\psi}^2 = 2,5333 = \frac{1}{4} \times 10,1333 = 2,5333 \text{ ، } \hat{\psi}^2 = 48,13$$

بدرجة حرية واحدة . وهذا العدد هو بالضبط العدد الذى وجدناه فى المثال (٨ - ٣) .

ويترتب على التعريف (٢) أن

$$\hat{\psi}^2 = 1 \div \hat{\psi}^2 = \hat{\psi}^2$$

وحين تتوفر الشروط المعتادة لتحليل التباين يكون توزيع نسبة التباين لأى مقارنة  $\hat{\psi}$  وهى

$$(29) \quad \frac{\hat{\psi}^2}{\epsilon^2} = F$$

مطابقا لتوزيع ف بدرجة حرية ١ ،  $\epsilon^2$  ويمكن باستخدام (٢٨) ، (٢٣) إثبات

$$\text{أن هذه النسبة هى بذاتها نسبة التباين (٢٧) وهى } F = \frac{\hat{\psi}^2}{\epsilon^2}$$

نبحث الآن فى المقارنات المستقلة ، وفى هذا البحث تلزمنا القاعدتين الآتيتين اللتين يمكن إثباتهما رياضيا .

قاعدة (١) :

إذا كانت  $\hat{\psi}_1$  و  $\hat{\psi}_2$  مقارنتين مستقلتين على نفس البيانات فإن :

$$(30) \quad \hat{\psi}_1^2 + \hat{\psi}_2^2 = (\hat{\psi}_1^2 + \hat{\psi}_2^2)$$

أى أن مجموع مربعات مقارنتين مستقلتين يساوى مجموع مربعات إحداهما مضافا إلى مجموع مربعات الأخرى . كما أن مجموع المربعات الناتج من ضم المقارنتين يكون له درجتان من درجات الحرية .



وتتد هذه القاعدة لأى عدد متبى من المقارنات المستقلة ولذلك نقول إن المقارنات المستقلة تجميعية additive

فمثلا ، فى المثال (٨ - ٤) الذى يتساءل عما إذا كان هناك فروق بين السكريات النقية (١) و(٢) و(٤) ، رأينا فى المثال (٨ - ١٠) أن هذا التساؤل يتضمن المقارنتين  $\hat{\psi}_1$  و  $\hat{\psi}_2$  . ونظرا لاستقلال هاتين المقارنتين يجب حسب القاعدة (١) أن يكون المجموع  $\hat{\psi}_1^2 + \hat{\psi}_2^2$  مساويا لمجموع المربعات الذى حصلنا عليه فى المثال (٨ - ٤) . بالحساب نجد ما يلى :

$$\hat{\psi}_1^2 = ١,١ ، \hat{\psi}_2^2 = \frac{١}{١٠} = \frac{١}{١٠} (١ + ١) = ٠,٢٠$$

$$\hat{\psi}_1^2 + \hat{\psi}_2^2 = (١,١) + (٠,٢٠) = ١,٣٠ = \text{بلرجه حرية واحدة}$$

$$\text{كذلك ، } \hat{\psi}_3^2 = ٠,٣٥ = \frac{١}{٢} ، \hat{\psi}_4^2 = ٠,١٥ = \frac{١}{٢} ، \hat{\psi}_5^2 = ١٩٠,٨٢$$

$$\text{إذن } \hat{\psi}_1^2 + \hat{\psi}_2^2 + \hat{\psi}_3^2 + \hat{\psi}_4^2 + \hat{\psi}_5^2 = ١٩٠,٨٢ + ١,٣٠ = ١٩٢,١٢$$

بلرجات حرية عددها ٢

وهذا بالضبط ما وجدناه بالمثال (٨ - ٤) وبذلك يتحقق أن

$$\hat{\psi}_1^2 + \hat{\psi}_2^2 + \hat{\psi}_3^2 + \hat{\psi}_4^2 + \hat{\psi}_5^2 = \hat{\psi}_1^2 + \hat{\psi}_2^2 + \hat{\psi}_3^2 + \hat{\psi}_4^2 + \hat{\psi}_5^2$$

قاعدة (٢) :

إذا كان بتجربة ما ك من أقسام المعالجة وكان هناك ك - ١ من المقارنات المستقلة  $\hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_{k-1}$  بين متوسطات العينات فإن :

$$\hat{\psi}_1^2 + \hat{\psi}_2^2 + \dots + \hat{\psi}_{k-1}^2 + \hat{\psi}_k^2 = \text{أى } [ \hat{\psi}_1^2 + \hat{\psi}_2^2 + \dots + \hat{\psi}_{k-1}^2 + \hat{\psi}_k^2 ] \quad (٣١)$$

(بين أقسام المعالجة)

وتتضمن هذه القاعدة أن مجموع مربعات أى مقارنة هو جزء من مجموع المربعات بين أقسام المعالجة ( بدرجة حرية واحدة ) ، كما تتضمن أنه لا يمكن أن يزيد عدد المقارنات المستقلة عن ك - ١ مقارنة . وهناك حرية كبيرة فى اختيار مجموعة من ك - ١ من المقارنات المستقلة ويتم هذا الاختيار بحسب التساؤلات التى تجرى التجربة للإجابة عنها ، فإذا بدأنا بمقارنة معينة يمكن دائما تكوين الك - ٢ من المقارنات المستقلة الأخرى .

فمثلا ، اعتبر المثال (٨ - ١) والأمثلة الثلاثة التابعة له .

عدد أقسام المعالجة = ك = ٥

من المثال (٨ - ١) وجدنا أن ٢ ٢ ( بين الأقسام ) = ١٠٧٧,٣٢ بدرجات حرية ٤ .

المقارنات  $\hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3, \hat{\psi}_4$  المدونة بالجدول (٨ - ١٤) هى مقارنات مستقلة عددها ٤ ( = ك - ١ ) .

٢ ٢  $\hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3, \hat{\psi}_4$  = ٨٣٢,٣٢ + ٤٨,١٣ + ٦,٠٥ + ١٩٠,٨٢

بدرجات حرية ٤ = ١٠٧٧,٣٢

= ٢ ٢ ( بين الأقسام )

وبذلك نتحقق القاعدة (٢) .

ملاحظة :

إن كل ما ذكر فى هذا البند عن المقارنات المستقلة بين المتوسطات ينطبق على المقارنات المستقلة بين مجاميع العينات ، مع ملاحظة أن مجموع مربعات مقارنة بين المجاميع هى  $\hat{\psi}^2 / \text{مجموع } \text{مجموع } \hat{\psi}^2$  حيث  $\hat{\psi}^2$  مقارنة بين المجاميع ،  $\text{مجموع } \hat{\psi}^2$  العينة ر .

مقال (٨ - ١٤) :

وجدت المشاهدات الآتية في تجربة ما :

جدول (٨ - ١٧)

الأقسام				
(٤)	(٣)	(٢)	(١)	
٣	٣	٢	٤	
٥	٥	٦	٦	
٤	٧	٣	٥	
٥	٧	٥		
٣	٨			
١٧ = $\Sigma$	٥	٥	٤	٣
٨١ = $\Sigma$	٢٠	٣٠	١٦	١٥
٤,٧٦ = $\Sigma$	٤	٦	٤	٥

(أولاً) أوجد  $\chi^2$  ( بين الأقسام ) ،

(ثانياً) حدد أوزاناً لكل من المقارنات القبيلة الثلاث الآتية بحيث تكون مستقلة ،

(أ) المجموعة (١) ضد المجموعة (٢) .

(ب) المجموعتان (١) و (٢) معا ضد المجموعة (٣)

(ج) المجموعات (١) و (٢) و (٣) ضد المجموعة (٤)

(ثالثا) أوجد قيمة كل مقارنة وقيمة مجموع مربعاتها ثم اختبر دلالتها عند المستوى ٠,٠٥، علما بأن  $\chi^2_{ع} = ٢,٤٦$  بدرجات حرية ١٣. اكتب جدول التباين بالتفصيل.

الحل :

$$(أولا) ٢٢ (بين الأقسام) = \frac{١٥}{٣} + \frac{١٦}{٤} + \frac{٢٠}{٥} + \frac{٢٠}{٥} - \frac{٨١}{١٧}$$

$$١٣,٠٥٨ = \text{بدرجات حرية } ٣$$

(ثانيا) لكي تكون المقارنات مستقلة يلزم توفر ما يلي :

(١) شرط المقارنة (١٨) وهو  $\mu_{أ} = \mu_{ب}$

(٢) شرط الاستقلال (٢١) وهو  $\mu_{أ} = \frac{\mu_{أب}}{\mu_{ب}}$

مع ملاحظة أن المجموعات مختلفة الأحجام .

بقليل من العمليات الحسابية نجد أن الأوزان المطلوبة يمكن أن تؤخذ بالقيم المسجلة بالجدول الآتي أو بأى مجموعة من القيم تتناسب معها . ويمكن أن يسير العمل كالآتي :

جدول (٨ - ١٨)

(١)	(٢)	(٣)	(٤)	المقارنة
				المتوسط
٥	٤	٦	٤	
١	١ -	٥	٥	١ : (١) ضد (٢)
٣	٤	٧ -	٥	٣ : (٢,١) ضد (٣)
٣	٤	٥	١٢ -	٣ : (١, ٢, ٣) ضد (٤)

نبدأ بأسهل المقارنات وهي  $\hat{\psi}_1$  ونضع لها الوزنين ١، -١ ثم نفرض أن أوزان المقارنة  $\hat{\psi}_1$  هي ١، ١، ١، ١. لتحقيق شرط المقارنة ينبغي أن يكون  $١ = \frac{1}{3} + \frac{1}{3} + \frac{1}{3}$ .  
ولتحقيق شرط استقلال  $\hat{\psi}_1$  و  $\hat{\psi}_2$  ينبغي أن يكون:

$$٠ = ٠ + \frac{1 \times 1 - 1}{4} + \frac{1 \times 1}{3}$$

مع ملاحظة أن ٣ هو حجم العينة الأولى، ٤ حجم العينة الثانية.

$$\text{ولذاً } \frac{1}{3} = \frac{1}{4} \text{ أى } ١ : ١ : ١ = ٣ : ٤$$

$$\text{فإذا أخذنا } ١ = ٣ \text{ فإن } ٤ = ١، ٤ = ١، ٧ = ١$$

وهذه هي الأعداد التي وضعت بالصف الثاني بالجدول.

بالمثل، إذا فرضنا أن أوزان المقارنة  $\hat{\psi}_2$  هي ١، ١، ١، ١، ١، ١ فإن  $٠ = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3}$ .

$$\text{لاستقلال } \hat{\psi}_1، \hat{\psi}_2 : \frac{١ \times 1 - 1}{4} + \frac{١ \times 1}{3} = ٠ \text{ ومنها } ١ : ١ : ١ : ١ : ١ = ٣ : ٤$$

$$\text{لاستقلال } \hat{\psi}_1، \hat{\psi}_2 : \frac{١ \times 3}{3} + \frac{١ \times 4}{4} + \frac{١ \times 7}{5} = ٠$$

$$\text{ومنها } ١ + ١ + \frac{٧}{5} = ٣ = ١، ٤ = ١ \text{ فإن } ١٢ = ٠، ١٢ = ٠$$

وهذه هي الأعداد التي وضعت بالصف الثالث من الجدول.

$$\text{(الثالث) } ١ = ١ \times ٠ + (١ - ٤) \times ١$$

$$١,٧١٤ = \frac{12}{7} = \frac{12}{11} / 1 = [ (1 - 1) \times \frac{1}{4} + 1 \times \frac{1}{3} ] / 1 = (\hat{\psi}_1)^2$$

$$\therefore \text{ ف } ١ > \frac{١,٧١٤}{٢,٤٦}$$

ليست ذات دلالة

كذلك

$$11- = 6 \times 7 - 4 \times 4 + 0 \times 3 = 1, \hat{1}$$

$$7,202 = 16 \frac{1}{2} / 121 = \left( \frac{13}{2} + \frac{1}{4} + \frac{1}{4} \right) / (11-) = 1, \hat{2}$$

$$\therefore \text{في } 2,928 = \frac{7,202}{2,46} \quad \text{ليست ذات دلالة عند المستوى } 0,05$$

$$\text{لأن في } 4,75 = [13, 1] 0,05$$

بالمثل

$$13 = 1, \hat{3} \quad 22 = 4,142 = \text{في } 1,684 \quad \text{ليست ذات دلالة}$$

$$\text{نلاحظ أن } 22 \text{ (المقارنات الثلاث)} = 4,142 + 7,202 + 1,714$$

$$13,058 = \text{حيث } 3 = 2$$

وهذا يساوي 22 (بين الأقسام) كما نتوقع .

جول التباين هو :

جول (8 - 19)

مصدر التباين	22	د ج	تقدير التباين	ف
بين الأقسام	13,058	3	4,353	1,77
الخطأ	31,980	13	1,714	1 >
			7,202	2,928
			4,142	1,684
			2,460	
الكل	45,038	16		

$$\text{بما أن } F_{0.05, 13, 3} = 5.90 ، F_{0.05, 13, 1} = 4.70$$

نقبل الفرض الصفري بأنه لا يوجد فروق بين أقسام المعالجة ، كما نقبل أن  $\psi = \psi_1 = \psi_2 = 0$

(٨ - ١٣) اختبار المقارنات البعدية :

نعلم أننا لا نجرى اختبارات المقارنات البعدية إلا إذا وجدنا من تحليل التباين أن هناك دلالة لعامل التجريب أوضححتها قيمة  $F$  . كما نعلم أنه في اختبار هذه المقارنات لا يمكننا أن تكون مستقلة كما هو الحال في المقارنات القبلية .

وقد قدمنا بالبند (٨ - ٥ - ٢) أسلوبا لاختبار المقارنات البعدية يعتمد على إيجاد قيمة حرجة لمجموع المربعات إذا تعداها مجموع المربعات بين قسمين أو مجموعة من الأقسام يكون الاختلاف بينها ذا دلالة وإلا فهو ليس كذلك . ويتميز هذا الأسلوب بالبساطة والعمومية ويمكن استخدامه لاختبار أى مقارنة دون اشتراط تساوى أحجام العينات ، وهو في الوقت نفسه غير حساس للانحرافات عن الاعتدالية وعدم تجانس التباينات . والصيغة التي استخدمناها لذلك هي المتباينة (٦) بالبند المذكور وهي :

$$F_{\alpha} \leq (K - 1) E^2_{\psi} \quad \text{ف} \quad \alpha = 0.05, 1 - \beta = 0.95 \quad (32)$$

وهذه المتباينة يمكن كتابتها بدلالة قيمة المقارنة  $\hat{\psi}$  والخطأ المعياري لها  $E_{\psi}$  كالآتي :

$$\hat{\psi}^2 \leq (K - 1) E^2_{\psi} \quad \text{ف} \quad \alpha = 0.05, 1 - \beta = 0.95 \quad (33)$$

حيث  $E^2_{\psi} = E^2_{\psi} \times \frac{1}{n}$  ،  $K$  عدد أقسام المعالجة ،  $n$  الحجم الكلي للعينات .

والعدد الذي بالطرف الأيسر من المتباينة (٣٣) هو القيمة الحرجة للقيمة  $\hat{\psi}^2$  . فإذا كانت  $\hat{\psi}^2$  مساوية أو أكبر منها يرفض الفرض الصفري  $\psi = 0$  عند المستوى  $\alpha$  وتكون هذه المقارنة هي إحدى العوامل التي تسببت في الدلالة العامة

لعامل التجريب ، أما إذا كانت  $\hat{\psi}^2$  أقل من القيمة الحرجة فإنها لا تكون ذات دلالة .

مثال (٨ - ١٥) :

في المثال (٨ - ٥) كان أحد التساؤلات البعدية هو : هل متوسط السكروز يختلف عن متوسط السكريات الثلاثة الأخرى مجتمعة ؟

هذا التساؤل يمكن وضعه على هيئة مقارنة كالتالي :

$$\hat{\psi}^2 = \frac{1}{1} \bar{S}_1 - \frac{1}{3} \bar{S}_2 + \frac{1}{3} \bar{S}_3 + \frac{1}{3} \bar{S}_4 =$$

$$= \frac{1}{3} (59,3 + 58,2 + 58) - 64,1 = 0,6$$

$$\hat{\psi}^2 = 31,36$$

$$\begin{aligned} &= 0,46 \times \left( \frac{1}{1} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} \right) = 0,728 \\ &= 0,1333 \times 0,46 = 0,0613 \end{aligned}$$

$$\therefore \text{القيمة الحرجة} = 4 \times 0,728 \times 2,58 = 7,513$$

بما أن  $31,36 < 7,513$  نرفض أن  $\psi = 0$  عند مستوى الدلالة ٠,٠٥

(٨ - ١٣ - ١) مقارنة أزواج المتوسطات :

تستخدم المتباينة (٣٣) للاختبارات البعدية لأزواج المتوسطات . ففي المثال (٨ - ١) كان لدينا خمسة أقسام واذن يكون هناك  $10 = \frac{5 \times 4}{2}$  اختبارات كل



منها على الصورة  $\bar{S}_1 - \bar{S}_2$  . في هذه الحالة يكون لدينا قيمة حرجة واحدة لجميع الاختبارات لأن قيمة  $\chi^2$  واحدة لأي مقارنة وهي :

$$\chi^2 = 0,46 \times \frac{1}{1} \times (1+1) = 0,92$$

وتكون القيمة الحرجة هي  $11,269 = 2,58 \times 0,92 \times 4$

ومن المناسب هنا وضع جميع المقارنات العشرة ( الفروق بين المتوسطات ) في جدول كالتالي حيث الأعداد داخل الجدول تعبر عن الفروق بين المتوسطات .

جدول ( ٨ - ٢٠ )

(٥)	(٤)	(٣)	(٢)	(١)	المتوسطات
٧٠,١	٦٤,١	٥٨,٠	٥٨,٢	٥٩,٣	
٩٠,٨-	٤,٨-	١,٣	١,١	٠	٥٩,٣ (١)
٩١,٩-	٥,٩-	٠,٢	٠		٥٨,٢ (٢)
٩٢,١-	٦,١-	٠			٥٨,٠ (٣)
٦,٠-	٠				٦٤,١ (٤)
٠					٧٠,١ (٥)

وبمضاهاة القيمة المطلقة لكل من هذه الفروق بالجذر التربيعي للقيمة الحرجة وهو

$\sqrt{11,269} = 3,357$  نجد أن سبعة منها ذات دلالة عند المستوى ٠,٠٥ وهي المشار إليها بنجمة في الجدول .

## تمارين (٨ - ٦)

(١) في احدى التجارب ذوات العامل الواحد كان هناك خمسة مستويات لعامل التجربة ، وقد اختير للتجريب خمس عينات عشوائية مستقلة بكل منها ١٠ مشاهدات وجد أن متوسطاتها كما هو مبين بالجدول الآتي :

المقارنات	المتوسطات				
	(١)	(٢)	(٣)	(٤)	(٥)
	٨٦	٩٥	٩٢	٨٠	١٠٤
$\hat{\mu}_1 - \hat{\mu}_2$	١	١ -	٠	٠	٠
$\hat{\mu}_1 - \hat{\mu}_3$	٠	٠	٠	١	١ -
$\hat{\mu}_1 - \hat{\mu}_4$	$-\frac{1}{4}$	$-\frac{1}{4}$	٠	$-\frac{1}{4}$	$-\frac{1}{4}$
$\hat{\mu}_1 - \hat{\mu}_5$	$-\frac{1}{4}$	$-\frac{1}{4}$	١ -	$-\frac{1}{4}$	$-\frac{1}{4}$

- (أولاً) اثبت أن المقارنات الملونة بالجدول مستقلة وأوجد قيمة كل منها .  
 (ثانياً) أوجد مجموع مربعات كل مقارنة واكتب جدول التباين بالتفصيل علماً بأن المجموع الكلي للمربعات ٥١١٢ . بين أن هناك دلالة لعامل التجربة ثم ابحث دلالة كل من المقارنات الأربعة .  
 (ثالثاً) أجر الاختبارات البعدية للفرق بين كل زوج من المتوسطات الخمسة .  
 (٢) أجريت تجربة لمعرفة العلاقة بين حجم ولون الحائط لحجرة تستخدم في أسلوب مقنن للمقابلات interviews وبين مستوى القلق للمختبرين ، وجاءت النتائج كما في الجدول الآتي :

اللون / الحجم	أزرق	أخضر	أصفر	أحمر	المجموع
صغير	٨٦	١٠٤	١٣٤	١٦٠	١٥٤٦
	٧١	١٧٥	١٣٩	١٥٥	
	١١٢	٩٦	١٤٤	١٧٠	
	٢٦٩	٣٧٥	٤١٧	٤٨٥	
متوسط	١١٠	٨٣	١٥٠	١٧٥	١٥٠٧
	٨٧	٨٩	١٥٦	١٥٢	
	١٠٠	٧٩	١٥٩	١٦٧	
	٢٩٧	٢٥١	٤٦٥	٤٩٤	
متوسط	١٠٥	٨٤	١٧٠	١٨٠	١٤٤٧
	٩٣	٨٦	١٣٣	١٥٤	
	٨٥	٨٣	١٢٨	١٤١	
	٢٨٣	٢٥٣	٤٣١	٤٧٥	
المجموع المتوسط	٨٤٩	٨٧٩	١٣١٣	١٤٥٤	٤٤٩٥
	٩٤,٣٣	٩٧,٦٧	١٤٥,٨٩	١٦١,٥٦	١٢٤,٨٦

( أولاً ) أجر تحليل التباين على أساس احتمال وجود تفاعل بين الخاصتين .  
( ثانياً ) أجر المقارنات القبلية الآتية بين الأعمدة ( الألوان ) واختبر دلالتها :  
 $\uparrow = \bar{S}_1 - \bar{S}_2 - \bar{S}_3 + \bar{S}_4$  ،  $\uparrow = \bar{S}_1 + \bar{S}_2 - \bar{S}_3 - \bar{S}_4$  ،  $\uparrow = \bar{S}_1 - \bar{S}_2 + \bar{S}_3 - \bar{S}_4$  ،  $\uparrow = \bar{S}_1 + \bar{S}_2 - \bar{S}_3 - \bar{S}_4$

( ثالثا ) أجر المقارنات البعدية بين جميع الفروق بين أزواج متوسطات الأعمدة ( الألوآن ) ،

( رابعا ) اختر مقارنتين مستقلتين بين الصفوف ( الأحجام ) واختبر دلالة كل منهما .

# ( ٨ - ١٤ ) النموذج عشوائى التأثيرات RANDOM EFFECTS MODEL

فى تناولنا لتحليل التباين للتجارب ذوات العامل الواحد بالبند ( ٨ - ٤ ) افترضنا أن لدينا مجتمعا معتدلا متوسطه  $\mu$  وتباينه  $\sigma^2$  أخذت منه عينة عشوائية قسمت عشوائيا إلى  $k$  من المجموعات لتتلقى  $k$  من المعالجات المختلفة مما قد يؤدى إلى اختلاف تراكيب هذه المجموعات ، ولذلك اعتبرنا أن هذه المجموعات هى عينات مأخوذة من مجتمعات (معتدلة) قد تختلف متوسطاتها  $\mu_1, \mu_2, \dots, \mu_k$  وإن كنا نعتبر أن لها تباين مشترك  $\sigma^2$  هو تباين المجتمع الأسمى . ونظرا لأن عناصر كل مجموعة تتلقى معالجة محددة فإننا نعتبر أن تأثير أى معالجة هو تأثير ثابت على جميع عناصر المجموعة التى تلقت هذه المعالجة ، ويختلف هذا التأثير من مجموعة إلى أخرى . ومن ثم وصفنا النموذج الإحصائى الذى استخدمناه فى التحليل بأنه نموذج ثابت التأثيرات أو النموذج II ووضعناه بالصيغة (٢) وهى :

$$S_{ijk} = \mu + \alpha_j + \gamma_{ij} \text{ حيث } \gamma_{11} = \gamma_{12} = \dots = \gamma_{1k} = 0, \dots, \gamma_{k1} = \gamma_{k2} = \dots = \gamma_{kk} = 0$$

وحيث  $S_{ijk}$  هو العنصر الرأى من المجموعة  $j$  التى تلقت المعالجة  $i$

$$\alpha_j = \mu_j - \mu \text{ ( الفرق بين متوسط المجتمع } j \text{ ومتوسط المجتمع$$

الأسمى ) وهو فرق ثابت بالنسبة لعناصر المجموعة } و

ويختلف من مجموعة إلى أخرى .

،  $\gamma_{ij}$  تعبر عن أخطاء التجريب وهو متغير عشوائى افترضنا أن له توزيع

معتدل : مع ( ٠ ،  $\sigma^2$  ) .

ويهدف هذا النموذج إلى المقارنة بين المتوسطات  $\mu_1, \mu_2, \dots, \mu_k$  .

في هذا النموذج ننظر إلى المعالجات التي استخدمت في التجريب على أنها تستغرق جميع المعالجات ذات الأهمية وتقتصر استنتاجاتنا فقط على هذه المعالجات المثلة في التجربة . ولكن هناك أنواع من التجارب يكون المطلوب فيها التوصل إلى استنتاجات عن مجموعة كبيرة من المعالجات تشمل المعالجات المثلة وغير المثلة في التجربة . أى أن الباحث يكون مهتما بمجموعة كبيرة من المعالجات الممكنة لعامل التجريب ولكنه حين يقوم بالتجربة لا يأخذها جميعها بل يأخذ عينة عشوائية منها ومع ذلك يرغب في التوصل إلى استنتاجات عن تأثيرات المجموعة الكاملة . في هذه الحالة يستخدم نموذجا إحصائيا آخر يسمى بالنموذج عشوائى التأثيرات أو بالنموذج II يصاغ كالآتي :

$$\begin{aligned} \text{حيث } m=1, 2, \dots, n_j \\ \text{حيث } \mu = \mu_1 + \mu_2 + \dots + \mu_m \\ \text{حيث } \mu = 1, 2, \dots, k \end{aligned} \quad (34)$$

$$\text{وحيث } \mu = \mu - (\mu)$$

= الفرق بين المتوسط  $\mu$  (ق) للمعالجة ق ومتوسط المجتمع الأصلي .  
ولما كان  $\mu$  (ق) هو متغير عشوائى لأن المعالجة ق تختار عشوائيا ، فإن  $\mu$  يكون بالضرورة متغيرا عشوائيا هو الآخر .

وهذا النموذج يشبه النموذج I إلا أن الفرق بينهما كبير ، فبينما نعتبر أن تأثيرات  $\alpha_j$  في النموذج II ثابتة ، فإننا نعتبر أن تأثيرات  $\mu_j$  عشوائية . أى أننا حين نكرر سحب العينات فإننا نحت النموذج I نسحب دائما نفس المعالجات بنفس التأثيرات  $\alpha_j$  ، أما تحت النموذج II فنسحب في كل مرة عينة عشوائية جديدة تختلف فيها تأثيرات  $\mu_j$  . ومن ثم نصف تأثيرات  $\mu_j$  بأنها عشوائية وتتوقف على العينة المختارة . وكما سبق القول يستخدم النموذج I حين يكون المطلوب التوصل إلى استنتاجات عن فئة محددة من المعالجات تستخدم جميعها في التجريب ، أما النموذج II فيستخدم حين تكون هناك فئة كبيرة من المعالجات التي تهتم الباحث ولكنه لا يستخدمها جميعا بل يستخدم عينة عشوائية منها .

وفي النموذج II نفترض أن للمتغيرين العشوائيين  $X$  و  $Y$  التوزيعين الآتيين :

$$X : \text{مع } (\sigma, 0) \text{ و } Y : \text{مع } (\sigma, 0) \quad (35)$$

كما نفترض أن قيم  $X$  ،  $Y$  مستقلة عن بعضها البعض .

وبلاحظ أننا عرفنا  $X$  بأنها انحرافات متوسطات المعالجات عن المتوسط العام للمجتمع ولذلك فإن متوسط تأثيرات  $X$  يساوى صفرا ، وهذا ما دعانا لأن نفترض أن متوسط توزيع  $X$  يساوى صفرا ، أما تبين هذا التوزيع فهو مقدار مجهول  $\sigma^2$  يراد تقديره وتقدير مدى مساهمته في الاختلاف الذي يظهر عند تحليل التباين وهذا أمر هام في كثير من التطبيقات الإحصائية .

إن تحليل التباين يتخذ نفس الأسلوب الحسابي في النموذجين I و II إلا أن الهدف من التحليل يختلف تماما ، فبينما يهدف التحليل في النموذج I إلى مقارنة المتوسطات ، فهو يهدف في النموذج II إلى دراسة التباينات ، وبصفة خاصة إلى تقدير واختبار كل من التباينين  $\sigma^2$  و  $\sigma^2$  وكذلك إلى تقدير التباين  $\sigma^2$  لتوزيع المعاينة للمتوسط الحسابي لأهميته في تقدير مدى الثقة في تقدير المتوسط الحقيقي للمجتمع عن طريق المتوسط  $\bar{y}$  للعينات. أما متوسطات تأثيرات  $X$  فلا جدوى من محاولة تقديرها أو تقدير الفروق بينها لأنها عشوائية .

إذا رمزنا بالرمز  $E$  للتباين المشاهد داخل أقسام المعالجة : ط ٢ ( داخل الأقسام ) ، وبالرمز  $E^1$  للتباين المشاهد بين الأقسام : ط ٢ ( بين الأقسام ) وبالرمز  $E^2$  لعدد قيم المتغير في أى قسم فيمكن إثبات ما يلي :

$$(1) E^2 \text{ هو تقدير غير متحيز للتباين } \sigma^2 \quad (36)$$

(٣٧)  $\sigma' \text{ هو تقدير غير متحيز للمقدار } \sigma + \sigma'$   
 وإذا رمزنا بالرمز  $\sigma'$  للتقدير غير المتحيز للتباين  $\sigma'$  فإن

$$(٣٨) \quad \sigma' = \frac{1}{n} (\sigma' - \sigma)$$

وهذه نتيجة مباشرة من (١) ، (٢) ، كما ينتج أن الفرض الصفري  $\sigma' = 0$  يمكن أن يختبر بواسطة اختبار ف بالصورة الآتية لأنه في حالة صحة هذا الفرض يكون كل من  $\sigma'$  ،  $\sigma'$  تقديرًا غير متحيز للتباين  $\sigma'$  :

$$(٣٩) \quad \sigma' = \frac{\sigma'}{\sigma} \text{ بدرجة حرية (ك - ١ ، هـ - ك)}$$

حيث ك عدد أقسام المعالجة و هـ = الحجم الكلي للعينات = ك هـ .

$$(٤٠) \quad (٣) \text{ تقدير } \sigma' \text{ هو } \frac{\sigma'}{\sigma}$$

أى أن التباين  $\sigma'$  لتوزيع المعاينة للمتوسط الحسابي يقدر بواسطة التباين المشاهد بين الأقسام مقسوما على العدد الكلي للملاحظات هـ .  
 تبين نوعية التجارب التى تستلزم النموذج II من المثالين الآتيين .

مثال (٨ - ١٦) :

في دراسة لمحتوى الكلسيوم في أوراق نبات اللفت الأخضر أخذت عينة عشوائية من ٤ أوراق من هذا النبات ثم أخذ من كل ورقة اختبرت ٤ أجزاء وزد كل منها ١٠٠ جرام وبذلك تجمع ١٦ جزءا من أوراق النبات . قيست النسبة المئوية لمحتوى الكلسيوم في كل منها وسجلت القياسات بالجدول (٨ - ٢١) الآتى :

جدول (٨ - ٢١)

النسب المئوية للكلسيوم في أوراق نبات اللفت : ك = ٤ أوراق ، ح = ٤ أجزاء .

	الأوراق			
	(١)	(٢)	(٣)	(٤)
النسبة المئوية للكلسيوم في أجزاء الورقة	٣,٢٨	٣,٥٢	٢,٨٨	٣,٣٤
	٣,٠٩	٣,٤٨	٢,٨٠	٣,٣٨
	٣,٠٣	٣,٣٨	٢,٨١	٣,٢٣
	٣,٠٣	٣,٣٨	٢,٧٦	٣,٧٦
المجموع للورقة	١٢,٤٣	١٣,٧٦	١١,٢٥	١٣,٢١
المتوسط للورقة	٣,١١	٣,٤٤	٢,٨١	٣,٣٠

نظرا لأن الأوراق تختلف من جوانب كثيرة قد لا نعرف طبيعتها كالاختلافات الوراثية والبيئية ونظرا لأننا اخترنا عينة عشوائية منها فإن النموذج المناسب هو النموذج عشوائى التأثير بالصيغة (٣٤) حيث  $\mu = 1, 2, 3, 4$  ترمز إلى نسبة الكلسيوم في الجزء ر من الورقة ،  $\mu = 1, 2, 3, 4$  حيث  $\mu$  ترمز إلى المتوسط الحقيقى لنسبة الكلسيوم في مجمع أوراق اللفت الأخضر . كما أن :

$\sigma^2$  ترمز إلى التباين بين الأوراق أى إلى مدى أثر الاختلافات الوراثية بين ورقة وأخرى على محتوى الكلسيوم ،



$\sigma^2$  ترمز إلى التباين داخل الأوراق أى بين القياسات داخل كل ورقة وبالتالى فهى تعبر عن تأثير الاختلافات غير الوراثية على محتوى الكلسيوم .

المطلوب فى هذا المثال ما يلى :

( أولا ) اختبار وجود أو عدم وجود تأثيرات ترجع إلى المعالجات أى إلى العوامل الوراثية . ويتضمن هذا الاختبار اختبار الفروض الصفرية :  $\sigma^2 = 0$  ،  $\sigma^2 = 0$  ،  $\sigma^2 = 0$  ،  $\sigma^2 = 0$  . وتحقق هذه الفروض إذا وقط إذا تحقق الفرض  $\sigma^2 = 0$  . ولذلك فإن الفرض الصفرى المطلوب اختباره يؤول إلى الآتى :

$$F : \sigma^2 = 0$$

وهذا الاختبار لا يتعلق فقط بوجود تأثيرات للمعالجات التى استخدمت فى التجربة بل لجميع المعالجات الممكنة التى دخلت والتى لم تدخل فيها .

( ثانيا ) تقدير المتوسط  $\mu$  لمحتوى الكلسيوم فى مجمع أوراق اللفت مع تقدير درجة الثقة فى هذا التقدير . وهذا هو الهدف الرئيسى فى هذه التجربة .

( ثالثا ) المقارنة بين التباينين  $\sigma^2$  و  $\sigma^2$  لأن هذه المقارنة قد تشير إلى ضرورة تعديل التجربة وتصميم تجربة مماثلة تكون أكثر دقة وأقل تكلفة . فإذا كان التباين  $\sigma^2$  بين الأوراق أكبر نسبيا من التباين  $\sigma^2$  داخل الأوراق فالأفضل تصميم تجربة نأخذ فيها عددا أكبر من الأوراق وعددا أصغر من الأجزاء والمكس بالعكس ، لأن هذا من شأنه تصغير تباين توزيع المعاينة للوسط الحسابى فيكون تقديرنا للبارامتر  $\mu$  أكثر دقة .

نقوم الآن بتحليل التباين للمثال ( ٨ - ١٦ ) .

$$\text{عامل التصحيح} = \frac{12}{2} = \frac{150.60}{16} = 160.375$$

$$١٦٠,٣٣٨٩ - '٣,٢٦ + \dots + '٣,٠٩ + '٣,٢٨ = (الكلي) ٢٢$$

$$١٥ \text{ بلوجات حرية} = ٠,٩٦٦٦$$

$$(١٣,٢١ + ١١,٢٥ + '١٣,٧٦ + '١٢,٤٣) - \frac{1}{4} = (بين الأوراق) ٢٢$$

$$١٦٠,٣٣٨٩ -$$

$$٣ \text{ بلوجات حرية} = ٠,٨٨٨٣٧$$

$$٠,٨٨٨٣٧ - ٠,٩٦٦٦ = (داخل الأوراق) ٢٢$$

$$١٢ \text{ بلوجات حرية} = ٠,٠٧٨٢٣$$

جدول (٨ - ٢٢)

مصدر التباين	٢٢	٣	١٥	التباين المتوقع
بين الأوراق	٠,٨٨٨٣٧	٣		'٤ - ٠,٢٩٦١
بين القياسات داخل الأوراق	٠,٠٧٨٢٣	١٢		'٤ - ٠,٠٠٦٦
الكل	٠,٩٦٦٦	١٥		'٤ + '٤ - '٤

(أولاً) : الفرض الصفري :  $'\sigma = ٠$  ( لا يوجد تأثير للخواص الوراثية على محتوى الكلسيوم )

$$من (٣٩) : ف = \frac{٠,٢٩٦١}{٠,٠٠٦٦} = ٤٤,٩^{**}$$

وهذه القيمة تزيد كثيرا عن القيمة الحرجة ف  $t_{12,0.01} = 0,95$  مما يجعلنا نرفض الفرض الصفري عند مستوى عالى من الدلالة ونستنتج أن هناك تفاوتاً كبيراً في الخواص الوراثية لأوراق النبات يؤثر تأثيراً فعالاً في محتوى الكلسيوم في هذه الأوراق .

( ثانياً ) : نقدر الوسط الحسابى  $\mu$  للنسبة المئوية لمحتوى الكلسيوم في مجتمع أوراق نبات اللفت الأخضر بواسطة الوسط الحسابى العام للعينات وهو  $3,17 =$  وليان مدى الدقة في هذا التقدير وحساب فترات الثقة للمتوسط  $\mu$  نستخدم الصيغة (٤٠) لتقدير تباين توزيع المعاينة للوسط الحسابى كالتالى :

$$\text{تقدير } \sigma^2 = \frac{s^2}{n} = \frac{0,2961}{16} = 0,0185 \quad \text{بدرجات حرية } 3$$

ويمكننا أن نوجد فترة ثقة للمتوسط  $\mu$  بدرجة ٩٩٪ كالتالى :

$$\text{الحد الأدنى للفترة} = 3,17 - 0,0185 \sqrt{t_{12,0.01} \times 16}$$

$$2,376 = 0,841 \times 0,1360 - 3,17 =$$

$$\text{الحد الأعلى للفترة} = 3,17 + 0,841 \times 0,1360 = 3,964$$

أى أن فترة الثقة المطلوبة هي ( ٢,٣٧٦ ، ٣,٩٦٤ ) .

( ثالثاً ) : من (٣٦) ، التقدير غير المتحيز للتباين  $\sigma^2$  هو  $s^2 = 0,0066$

ومن (٣٨) ، التقدير غير المتحيز للتباين  $\sigma^2$  هو

$$s^2 = \frac{(s^2 - s^2)}{n-1} = \frac{(0,0066 - 0,2961)}{15} = 0,0724$$

$$\text{نلاحظ أن } \frac{s^2}{s^2} = \frac{0,0724}{0,0066} = 10,97$$

أى أن تقدير  $\sigma^2$  عشرة أضعاف أو أكثر تقدير  $\sigma^2$  ولو أردنا تحسين التجربة وتحسين الخطأ المعيارى لتوزيع المعاينة للوسط الحسابى ينبغى أن نزيد من عدد الأوراق وأن نقلل من عدد الأجزاء التى تؤخذ من كل ورقة .

#### ملاحظة :

فى هذا المثال اخترنا عينة عشوائية من الوحدات ( أوراق النبات ) التى يمكن أن نسميها بالوحدات الابتدائية للمعاينة *primary sampling units* ثم اخترنا عينات عشوائية جزئية من كل وحدة من الوحدات الابتدائية ويمكن أن نستمر فى ذلك بأخذ عينات عشوائية من كل عنصر من عناصر العينات الجزئية السابق اختيارها وأن نكرر ذلك حسبما تقتضى التجربة . إن مثل هذه المعاينة تسمى بالمعاينة ذات المراحل أو بالمعاينة العشبية *multistage or nested sampling* حيث تحدث عدة تقسيمات متدرجة ومتداخلة كأعشاش الطيور . وسوف نعود إلى ذلك فى البند (١٥ - ٥) من الفصل الخامس عشر .

#### مثال (٨ - ١٧) :

فى إحدى التجارب النفسية كان يشك فى أن شخصية المجرى ( القائم بالتجريب ) لها تأثير على النتائج التى يتوصل إليها . ونظرا لأن هناك عددا كبيرا من المجرىين الذين يمكنهم القيام بالتجربة مما يعوق استخدامهم جميعا فقد اختر منهم عينة عشوائية من خمسة مجرىين لإجراء التجربة تحت نفس الظروف على أن يستخدم كل منهم مجموعة من ٨ أشخاص تختار عشوائيا وعلى أن توزع المجموعات على المجرىين عشوائيا . سجلت البيانات الناتجة من التجارب الخمس فى الجدول (٨ - ٢٣) الآتى ، والمطلوب بحث ما إذا كان لشخصيات المجرىين أثر على نتائج التجربة .

الجدول ( ٨ - ٢٣ )

المجموع				
(١)	(٢)	(٣)	(٤)	(٥)
٥,٨	٦,٠	٦,٣	٦,٤	٥,٧
٥,٩	٦,١	٥,٥	٦,٤	٥,٩
٥,٧	٦,٦	٥,٧	٦,٥	٦,٥
٥,٩	٦,٥	٦,٠	٦,١	٦,٣
٥,٦	٥,٩	٦,١	٦,٦	٦,٢
٥,٤	٥,٩	٦,٢	٥,٩	٦,٤
٥,٣	٦,٤	٥,٨	٦,٧	٦,٠
٥,٢	٦,٣	٥,٦	٦,٠	٦,٣
٤٤,٠	٥٠,٦	٤٧,٢	٥٠,٦	٤٩,٣
٢٤٠,٨				

الحل :

نظرا لأن هناك عددا كبيرا من الأفراد الذين يمكنهم القيام بالتجربة ، لكل منهم شخصية تميزه عن الآخرين ، ونظرا لأننا اخترنا عينة عشوائية منهم ، فإن النموذج المناسب لهذه التجربة هو النموذج عشوائى التأثير بالصيغة (٣٤) ، ونعتبر أن لدينا خمس معالجات يمثلها خمسة مجريين .

$$\text{عامل التصحيح} = \frac{٢٤٠,٨^2}{٤٠} = ١٤٤٩,٦١٦$$

$${}^{\circ}6,3 + {}^{\circ}6,0 + \dots + {}^{\circ}5,9 + {}^{\circ}5,8 = \text{٢٢ (الكل)}$$

$$- 1499,616$$

$$\text{بلدجات حرية ٣٩} \quad 6,32 =$$

$$\text{٢٢ (بين المجريين)} = \frac{1}{8} ({}^{\circ}44,0 + {}^{\circ}50,6 + \dots + {}^{\circ}49,3) - 1449,616$$

$$\text{بلدجات حرية ٤} \quad 3,47 =$$

$$\text{٢٢ (داخل المجريين)} = 3,47 - 6,32 = 2,85 \quad \text{بلدجات حرية ٣٥}$$

نلخص هذه النتائج في جدول التباين الآتي :

جدول (٨ - ٢٤)

مصدر التباين	٢٢	ح د	ط	التباين الموقوع	في
بين المجريين	3,47	٤	$\text{ع}^{\circ} - \text{ح}^{\circ} = 0,868$	$\text{ع}^{\circ} + \text{ح}^{\circ} = 1,08$	$10,72^{\circ}$
داخل المجريين	2,85	35	$\text{ع}^{\circ} - \text{ح}^{\circ} = 0,081$	$\text{ع}^{\circ}$	
الكل	6,32	39			

الفرض الصفري :  $\sigma^{\circ} = 0$  (لا يوجد تأثير للمجريين على نتائج التجربة)

$$\text{في} = \frac{0,868}{0,081} = 10,72^{\circ}$$

وهذه القيمة تزيد عن القيمة الحرجة ف  $[30,4]_{0,01}$  التي تقع بين ٣,٨٣ ، ٤,٠٢ وإذن نرفض الفرض الصفري عند مستوى الدلالة ٠,٠١ ونحكم بأن هناك دليلا كافيا على صحة القول بأن لشخصيات المجرمين تأثيرا فيما يتوصلون إليه من نتائج في هذه التجربة . وهذه نتيجة خطيرة ينبغي أن تؤخذ في الاعتبار عند تفسير نتائج التجربة . وتبين هذه الخطورة أيضا إذا حسبنا النسبة التي ساهم بها التباين  $\sigma^2$  من التباين الكلي ( وهو  $6,32 \div 39 = 0,179$  ) :

من (٣٨) : التقدير غير المتحيز للتباين  $\sigma^2$  هو

$$E = \frac{1}{n} = (0,868 - 0,081) = 0,098$$

∴ النسبة التي ساهم بها التباين  $\sigma^2$  من التباين الكلي  $= \frac{0,098}{0,179} = 0,55$  وهي

نسبة كبيرة تشير إلى أن أكثر من نصف التباين الكلي يرجع إلى تأثير اختلاف شخصيات المجرمين .

**ملاحظة :**

يتم استخدام النموذج عشوائى التأثيرات لتحليل التباين للتجارب ذوات العاملين أو أكثر .





## الفصل التاسع

### الانحدار الخطى البسيط

#### SIMPLE LINEAR REGRESSION

##### (٩ - ١) المجتمع ذو المتغيرين : BIVARIATE POPULATION

المجتمع ذو المتغيرين هو مجتمع ننظر إليه من حيث متغيرين  $x$  ،  $y$  ،  $z$  ، يكونان موضع اهتمامنا في وقت ما ، فمثلا في مجتمع من الطلاب قد نهم بالمتغير  $x$  الذى يعبر عن نسبة الذكاء والمتغير  $y$  الذى يعبر عن درجة الطالب في مادة الإحصاء وفي مجتمع من القمح قد تعبر  $x$  عن تاريخ الزراعة وتعبر  $y$  عن مقدار المحصول الناتج ، وفي مجتمع من الأبقار قد تعبر  $x$  عن مقدار الغذاء اليومي وتعبر  $y$  عن الزيادة في الوزن بعد مدة من الزمن .

كما سبق الذكر ، يميز الإحصاء بين نوعين من المتغيرات : المتغيرات العشوائية والمتغيرات غير العشوائية ، والمتغير العشوائي هو متغير حقيقي يخضع لمؤثرات عشوائية غير منظورة ولذلك لا نستطيع التحكم فيه تجريبياً . وهذا النوع من المتغير يكون له توزيع احتمال بمعنى أنه لو كان من النوع الوثاب مثلا فإن ظهور أى عنصر من عناصره يكون مصحوباً باحتمال ما ، أما المتغير غير العشوائي ( أو الرياضي ) فليس له توزيع احتمال ويمكن التحكم فيه تجريبياً أو تحديد قيمه مقدماً أو قياسها بدقة أو بخطأ يمكن إهماله .

وفي دراستنا لمجتمع  $z$  متغيرين  $x$  ،  $y$  ، كثيراً ما يكون اهتمامنا منصباً على إيجاد العلاقة بينهما إن كان هناك ثمة علاقة ، وعلى محاولة وضع هذه العلاقة على

هيئة دالة  $\alpha = \beta + \gamma$  . وتختلف هذه الدالة بطبيعة الحال باختلاف العلاقة بين المتغيرين فقد تكون على صورة خطية  $\alpha = \beta + \gamma$  أو على صورة تربيعية  $\alpha = \beta^2 + \gamma$  أو على صورة أسية  $\alpha = e^{\beta} \gamma$  وهكذا ... إلا أننا سوف نقصر اهتمامنا هنا على الحالات التي تكون فيها الدالة على الصورة الخطية .

سنبنى هذه الدراسة على الافتراضين الآتيين وافترض ثالث تقدمه في البند (٩ - ٥) ، والنموذج الذى نستخدمه فى هذه الدراسة يسمى بالنموذج II لتحليل الانحدار .

### الافتراض الأول :

١ المتغير  $\alpha$  هو متغير رياضي والمتغير  $\beta$  هو متغير عشوائي .

وهذا الافتراض يعنى من الناحية التطبيقية أننا نبدأ بتحديد قيم ثابتة  $\alpha_1, \alpha_2, \dots, \alpha_n$  من المتغير  $\alpha$  ثم نقوم بملاحظة أو قياس القيم ( العشوائية ) المناظرة  $\beta_1, \beta_2, \dots, \beta_n$  من المتغير  $\beta$  . فمثلاً قد تكون القيم السينية  $\alpha$  هي درجات حرارة معينة ١٠ ، ١٥ ، ٢٠ ، ٣٠ ، ٤٠ ، ٥٠ وتكون القيم الصادية  $\beta$  هي مقادير ما نشاهده من تمدد معدن عند هذه الدرجات . أو تكون قيم  $\alpha$  هي أطوال أطفال عند الولادة بالاستمرات وتكون قيم  $\beta$  هي مدد الحمل بالأيام . أو تكون قيم  $\alpha$  هي أعماق محددة تحت سطح البحر وتكون القيم الصادية  $\beta$  هي نسب الملوحة عند هذه الأعماق .

### الافتراض الثاني :

٢ العلاقة بين المتغيرين  $\alpha$  ،  $\beta$  ( في المجمع ) هي علاقة خطية .

ونعني بهذا الفرض رياضياً أن يكون متوسط قيمة  $\beta$  عند قيمة معينة  $\alpha$  يأخذ الصيغة الخطية الآتية :

$$\beta = \alpha + \epsilon \quad (1)$$

حيث  $\alpha$  ،  $\beta$  بارامتران مجهولان ينبغي تقديرهما من العينة . هذا مع ملاحظة أنه عند أى قيمة  $s$  يمكن أن تأخذ  $s$  قيمة كثيرة لأنها كما ذكرنا تتأثر بعوامل عشوائية لا نستطيع التحكم فيها ولهذا يعبر الطرف الأيمن من الصيغة (١) عن متوسط هذه القيم وهو بالطبع أحسن قيمة تناظر القيمة  $s$  . ويسمى المتغير  $s$  بالمتغير المستقل كما يسمى المتغير  $s$  بالمتغير التابع .

إن هذا الافتراض لا يوضع إلا في الحالات التي نعلم بخيرتنا السابقة أو من معلومات خاصة أنه صحيح . ومع ذلك إذا كنا نشك في صحته فهناك طريقة إحصائية سنذكرها بعد للتحقق من سلامته .

إن مهمتنا الابتدائية في دراسة العلاقة الخطية بين متغيرين  $s$  ،  $s$  هي إيجاد أحسن تقديرين للبارامترين المجهولين  $\alpha$  ،  $\beta$  الموجودين بالعلاقة (١) ، ونعتمد في ذلك كالاعتداع على تصميم تجربة على عينة من المجتمع الذى ندرسه ومنها نحصل على قيم تجريبية للمتغيرين  $s$  ،  $s$  تبدو كما يلي :

$$\begin{array}{ccccccc} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 \\ s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 \end{array}$$

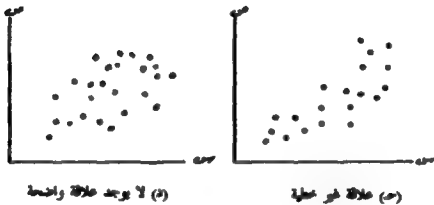
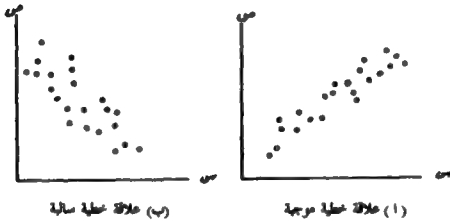
ومن هذه الأزواج من القيم نوجد التقديرين المطلوبين كما سنرى بعد ، ويساعدنا التمثيل البياني لهذه الأزواج على تصور المشكلة التي نتناولها والخط الذى نشده .

## SCATTERGRAM

(٩ - ٢) شكل الانتشار :

على أساس أن المتغيرين  $s$  ،  $s$  حقيقيان نستطيع تمثيل أزواج قيم  $(s, s)$  التى حصلنا عنها من العينة على نظام إحداثيات دى محورين متعامدين . فيمثل كل زوج منها بنقطة معينة في المستوى ، ومجموعة النقاط الناتجة تؤلف شكلاً يسمى بشكل الانتشار . وإذا كان هناك ثمة علاقة بين المتغيرين فإن هذه النقاط لا تكون مبعثرة كيفما اتفق بل تتخذ نمطاً معيناً يوحى بوجود وطبيعة هذه

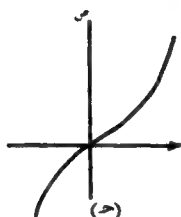
العلاقة . فإذا بدا شكل الانتشار كما في أى من الشكلين (٩ - ١ - ١) أو (٩ - ١ - ٢) فإنه يشير بصفة مبدئية بأن العلاقة بين المتغيرين هي علاقة خطية ، لأننا نستطيع أن نتصور وجود خط مستقيم تقع النقاط من حوله وقرية منه وإن كانت لا تمر جميعها به .



الشكل (٩ - ١) بعض صور شكل الانتشار

أما إذا بدا شكل الانتشار كما في أى من الشكلين (٣) ، (٤) فإننا نشك في خطية العلاقة بين المتغيرين وعلينا أن نبحث عن افتراض آخر غير افتراض الخطية نتعامل به مع المتغيرين . إن معرفة الباحث بخواص المنحنيات تساعد على كشف طبيعة

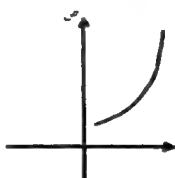
العلاقة بين المتغيرين واقتراح المعادلة التي تناسبها . وفيما يلي بعض المنحنيات التي تعاون على كشف الأنماط التي قد يشير إليها شكل الانتشار والدوال التي تعبر عنها جبريا . وسنرى في البند (٩ - ٧) أنه يمكن تناول بعض هذه الدوال كدوال خطية بعد إجراء تحويلات مناسبة عليها .



(أ)

$$y = \alpha x$$

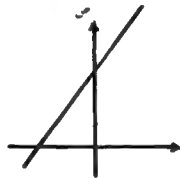
دالة تكهنية



(ب)

$$y = \alpha x^2 + \beta x + \gamma$$

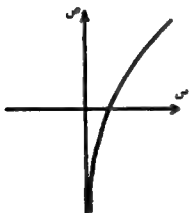
دالة تربيعية



(١)

$$y = \alpha x + \beta$$

دالة خطية



$$y = \alpha \log x + \beta$$

دالة لوغاريتمية



$$y = \alpha x^\beta + \gamma$$

دالة أسية



$$y = \frac{\beta}{x} + \alpha$$

دالة عكس متناسبة

الشكل (٩ - ٧) : بعض أنماط شكل الانتشار

### (٩ - ٣) تقدير البارامترين $\alpha$ ، $\beta$ .

نفرض أن  $(س_١ ، ص_١)$  هى إحدى القيم المشاهدة في العينة . حسب الافتراض الثاني ينبغي أن تكون العلاقة بين  $س_١$  ،  $ص_١$  على الصورة :

$$(٢) \quad ص_١ = \alpha + \beta س_١ + خ_١$$

حيث  $خ_١$  هو الفرق بين القيمة المشاهدة  $ص_١$  المناظرة للقيمة  $س_١$  وبين القيمة المتوقعة لها وهي  $\alpha + \beta س_١$  ولذلك تسمى القيم  $خ_١$  بالبواقي residuals وهي تعبر عن الأخطاء العشوائية في قياس المتغير العشوائي  $ص$  . ونظراً لأن هذه الأخطاء تكون بالزيادة لبعض قيم  $ص$  وبالنقصان للبعض الآخر ، فإننا نفترض أن متوسط هذه الأخطاء يؤول إلى الصفر على المدى البعيد .

### طريقة المربعات الصغرى : LEAST SQUARES METHOD

كما أشرنا من قبل ، إن مهمتنا الأساسية هى استخدام أزواج القيم  $(س_١ ، ص_١)$  المشاهدة في العينة لإيجاد أحسن تقديرين للبارامترين  $\alpha$  ،  $\beta$  . فإذا كان  $ا$  ،  $ب$  هما هذان التقديران فإن المعادلة :

$$(٣) \quad ص = ا + ب س$$

تكون معبرة أحسن تعبير عن المعادلة (١) ويكون الخط الممثل للمعادلة (٣) هو أحسن خط يلائم مجموعة نقاط العينة . فإذا وقفنا في إيجاد  $ا$  ،  $ب$  سمى الخط (٣) بخط أحسن مطابقة line of best fit أو خط انحدار  $ص$  على  $س$  regression line of  $Y$  on  $X$  وأحياناً يسمى بصيغة التنبؤ prediction formula .

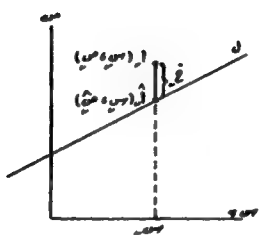
والطريقة الأكثر شيوعاً لإيجاد  $ا$  ،  $ب$  تبني على مبدأ يعرف بمبدأ المربعات الصغرى يمكن صياغته كآلاتي ، مع ملاحظة الافتراض الأول عن أن قيم  $س$  ثابتة وقيم  $ص$  متغيرة .

و أحسن خط يلام مجموعة النقط  $(س_١, ص_١)$  هو ذلك الذى نحدده بحيث يكون مجموع مربعات البواقي  $خ$  أصغر ما يمكن .

من المعادلة (٢) نرى أن  $خ = ص_١ - \alpha - \beta س_١$

ليكن  $د (\alpha, \beta) = خ' = خ'_١ = (ص_١ - \alpha - \beta س_١)$  (٤)  
 فإذا نظرنا للدالة (٤) على أنها دالة في المجهولين  $\alpha, \beta$  فإن مبدأ المربعات الصغرى يقول إن أحسن خط هو ذلك الذى نحدد فيه قيمتاً  $\alpha, \beta$  بحيث تكون قيمة هذه الدالة نهاية صغرى .

وبوضح هذا المبدأ هندسياً كالآتي :



الشكل (٩ - ٣)

لتكن  $أ' (س_١, ص_١)$  إحدى النقط المشاهدة في العينة أى إحدى نقط شكل الانتشار ، ولتكن  $أ' (س_١, ص_١)$  هي نقطة واقعة على خط مستقيم ل وتشارك مع النقطة  $أ' (س_١, ص_١)$  في الإحداثي السيني  $س_١$  .  
 إن مبدأ المربعات الصغرى يقول إن الخط ل يكون هو خط أحسن مطابقة إذا كان مجموع مربعات المسافات الرأسية بين النقط  $أ' (س_١, ص_١)$  نهاية صغرى ، أى إذا كانت الدالة

$$د (\alpha, \beta) = (ص_١ - \alpha - \beta س_١)^2 = (ص_١ - \alpha - \beta س_١)^2$$

نهاية صغرى . بطرق التفاضل المعتادة نفاضل الدالة (٤) جزئياً بالنسبة إلى كل من  $\alpha, \beta$  . وبمساواة كل من الناتجين بالصفر نحصل على معادلتين آتيتين يكون

حللها معاً هما القيمتان  $\alpha$  ،  $\beta$  المطلوبتان. وهاتان المعادلتان تسميان بالمعادلتين المعتادتين وتأخذان الصورتين الآتيتين :

$$\beta + \alpha = \bar{y} , \quad \beta + \alpha = \bar{y} = \bar{y} - \bar{x} = \bar{y} - \bar{x}$$

وبحللها معاً ينتج ما يلي :

أحسن تقديرين للبارامترين  $\alpha$  ،  $\beta$  من العينة هما  $\alpha$  ،  $\beta$  حيث :

$$(5) \quad \beta = \frac{n \bar{y} \bar{x} - (\sum x)(\sum y)}{n \bar{x}^2 - (\sum x)^2}$$

وحيث  $n$  هي عدد أزواج القيم  $(x, y)$

$$(6) \quad \alpha = \bar{y} - \beta \bar{x}$$

وبالتالي تكون معادلة الانحدار  $\hat{y}$  على  $x$  هي

$$\hat{y} = \bar{y} + \beta (\bar{x} - \bar{x})$$

$$(7) \quad \text{أو } \hat{y} = \bar{y} + \beta (\bar{x} - \bar{x})$$

حيث  $\beta$  معطاة بالصيغة (5) وتسمى بمعامل الانحدار  $\hat{y}$  على  $x$  وهي تعبر عن ميل خط الانحدار (7) عن الاتجاه الموجب لمحور السينات .

ملاحظة (1) :

من الواضح أن خط الانحدار (7) يمر بالنقطة  $(\bar{x}, \bar{y})$  .

ملاحظة (2) :

بقسمة كل من بسط ومقام الطرف الأيسر من (5) على  $(1 - r)$  وبعد عمليات جبرية بسيطة نجد أن :



$$\frac{(1) \text{ تغاير (س، ص) }}{\text{تباین س}} = \frac{\text{ع س ص}}{\text{ع س}} = \frac{1 - \frac{1}{2} \times (\text{س} - \text{ص})}{1 - \frac{1}{2} \times (\text{س} - \text{ص})} = \text{ب}$$

وفي إيجاد ب من (٥) أو (٨) يمكن أن نجتمع أو نطرح أى عدد من جميع قيم ب، وأى عدد من جميع قيم ص دون أن تتأثر قيمة ب. وفي حالة التوزيعات التكرارية حيث يكون لكل زوج (س، ص) تكرار له نحصل على ب من أى من الصيغتين بوضع له بعد كل رمز ص.

مثال (۹ - ۱) :

أوجد معادلة انحدار  $y$  على  $x$  من البيانات الآتية :

٥,٠ ٤,٨ ٤,٤ ٣,٩ ٣,٥ ٣,٠ ٢,٤ ١,٨ ١,٥ : س

ومنها أوجد أحسن تقدير لقيمة  $s$  عندما  $s = 3,7$ .

**الحل :**

( يلاحظ أننا لو رسمنا شكل الانتشار لوجدنا أن افتراض الخطية هو افتراض معقول ) . يتطلب الحل إيجاد كل من  $x_1$  ،  $x_2$  ،  $y_1$  ،  $y_2$  ،  $s_1$  ،  $s_2$  ، و هذه جميعاً نحصل عليها من الجدول ( ٩ - ١ ) الآتي .

لاحظ أن العمود الأخير لا لزوم له في إيجاد معادلة الانحدار ص على س، ولكنه سيزمننا فيما بعد في تحليل الانحدار. ويمكن استخدام حاسبات الجيب للحصول على هذه المجاميع دون ضرورة لكتابة التفاصيل المبينة بالجداول وذلك توفيراً للوقت والجهد.

جدول (٩ - ١)

إيجاد معادلة خط الانحدار من بيانات المثال (٩ - ١)

س	ص	س ص	س <sup>٢</sup>	ص <sup>٢</sup>
١,٥	٤,٨	٧,٢٠	٢,٢٥	٢٣,٠٤
١,٨	٥,٧	١٠,٢٦	٣,٢٤	٣٢,٤٩
٢,٤	٧,٠	١٦,٨٠	٥,٧٦	٤٩,٠٠
٣,٠	٨,٣	٢٤,٩٠	٩,٠٠	٦٨,٨٩
٣,٥	١٠,٩	٣٨,١٥	١٢,٢٥	١١٨,٨١
٣,٩	١٢,٤	٤٨,٣٦	١٥,٢١	١٥٣,٧٦
٤,٤	١٣,١	٥٧,٦٤	١٩,٣٦	١٧١,٦١
٤,٨	١٣,٦	٦٥,٢٨	٢٣,٠٤	١٨٤,٩٦
٥,٠	١٥,٣	٧٦,٥٠	٢٥,٠٠	٢٣٤,٠٩
٣٠,٣	٩١,١	٣٤٥,٠٩	١١٥,١١	١٠٣٦,٦٥

$$\text{من الصيغة (٥) نجد أن ب} = \frac{٩١,١ \times ٣٠,٣ - ٣٤٥,٠٩ \times ٩}{(٣٠,٣) - ١١٥,١١ \times ٩} = ٢,٩٣٠٣$$

$$\text{كما أن س} = \frac{٣٠,٣}{٩} = ٣,٣٦٦٧, \quad \text{ص} = \frac{٩١,١}{٩} = ١٠,١٢٢٢$$

∴ معادلة انحدار ص على س (من الصيغة ٧) هي

$$\text{ص} = ٢,٩٣٠٣ (\text{س} - ٣,٣٦٦٧) + ١٠,١٢٢٢$$

ومنها  $ش = ٠,٢٥٦٨ + ٢,٩٣٠٣ س$

وحينما  $س = ٣,٧$  فإن أحسن تقدير لقيمة  $ص$  هي :

$$ش = ٠,٢٥٦٨ + ٢,٩٣٠٣ \times ٣,٧ = ١١,٠٩٨٩$$

#### (٩ - ٤) الخطأ المعياري للتقدير Standard Error of Estimate

كما هو الحال عند دراسة بيانات عن متغير واحد حيث نصف هذه البيانات بواسطة الوسط الحسابي الذي يعطى تقديراً لمتوسط هذه البيانات ثم ندعم هذا الوصف بتقدير التشتت بواسطة الانحراف المعياري ، فاننا نصف البيانات ذوات المتغيرين بواسطة خط الانحدار الذي يعطى تقديراً لمتوسطات قيم  $ص$  عند قيم  $س$  ونستكمل هذا الوصف بتقدير مدى تشتت نقط شكل الانتشار حول هذا الخط . ويقاس هذا التشتت بما يسمى بالخطأ المعياري لتقدير  $ص$  من معادلة الانحدار (٧) ، ويرمز له بالرمز  $ع.م.س$  ويعرف كآتي :

$$ع.م.س = \frac{1}{n-2} (ص.س - ص.ش) \quad (٩)$$

وهذا المقياس له أهمية كبرى في عمليات الاستنتاج الإحصائي كما سنرى بعد .  
ملاحظة :

يمكن إثبات أن .

$$ع.م.س = \frac{1}{n-2} (ص.س - ص.ش - ص.ب) \quad (١٠)$$

وهذه الصيغة أسهل في حساب الخطأ المعياري من الصيغة (٩) وتستخرج من نفس المجاميع التي نوجد لها حساب معادلة الانحدار .

مثال (٩ - ٢) :

أوجد الخطأ المعياري لخط الانحدار الناتج في المثال (٩ - ١) السابق .

الحل :

نعوض في الصيغة (١٠) من مجاميع الجدول (٩ - ١) بالقيمتين

$$1 = 0,2068 , 2 = 2,9303$$

$$ع_{م.س} = \frac{1}{4} (1036,06 - 91,1 \times 0,2068 - 2,9303 \times 245,09)$$

$$= 0,2912$$

$$\therefore ع_{م.س} = 0,0396 \text{ (الخطأ المعياري مقدراً من العينة)}$$

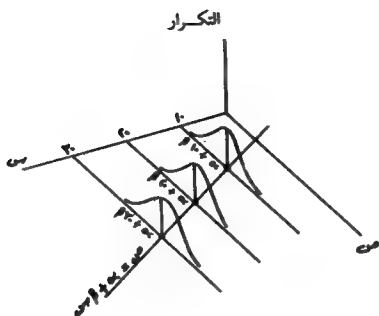
(٩ - ٥) استنتاجات إحصائية :

إن إيجاد معادلة الانحدار الخطي لم يستلزم إلا وضع الافتراضين الأول والثاني السابق ذكرهما ، أما إذا أردنا القيام باستنتاجات إحصائية أى إصدار قرارات عن المجتمع عن طريق العينة فإن ذلك يتطلب وضع افتراض خاص عن توزيع احتمال المتغير العشوائي  $\mu$  . وفي المعتاد نضع الافتراض الآتي (الذي يمكن هو أيضاً اختبار صحته) .

الافتراض الثالث :

« عدد أى قيمة ثابتة  $\mu$  يكون للمتغير العشوائي  $\mu$  توزيع معدل متوسطه  $\alpha + \beta \mu$  وتباينه عدد ثابت مجهول  $\sigma^2$  مستقل عن  $\mu$  » .

ويمكن تصوير هذا الافتراض كالآتي :



الشكل (٩ - ٤) توزيعات معطلة للمطفر مع عدد  $n = 10, 20, 30$

وبلاحظ ما يلي :

(أ) متوسطات التوزيعات الصادية تختلف باختلاف قيم  $n$  وهى تقع جميعها على خط الانحدار .

(ب) جميع التوزيعات الصادية متوازية ولها نفس الشكل لأن لها نفس التباين وتختلف فقط في مواقعها كما جاء في (أ) .

من الافتراضات الثلاثة المذكورة ، بالإضافة إلى افتراض العشوائية الذى يضمن استقلال وحدات التجريب عن بعضها البعض ، يمكن أن نخرج بعدة استنتاجات نختار منها ما يلي :

( أولا ) اختبار الفرض  $\beta = 1$  حيث  $k$  عدد معين .

للوصول إلى خط الانحدار  $\hat{y} = 1 + b$  نوجد معامل الانحدار  $b$  من واقع بيانات مأخوذة من عينة ما ، وإذا اخترنا عينة أخرى نحصل على قيمة مختلفة

لهذا العامل . أى أن قيمة ب تختلف من عينة إلى أخرى ، ولذلك نعتبر أن أى قيمة للمعامل ب هي إحدى القيم المشاهدة من متغير عشوائى B ذى عدد غير متنى من القيم . ونحت الافتراضات الثلاثة للانحدار يمكن إثبات أن لهذا المتغير توزيعا معتدلا وسطه الحسابى  $\beta$  وانحرافه المعياري يقدر بالمقدار  $\sigma_{\beta} = \sigma / \sqrt{22}$  حيث  $\sigma$  هو الخطأ المعياري للتقدير المعطى بالصيغة (٩) وحيث

$$\sigma^2 = \frac{1}{n-2} \sum (y_i - \hat{y}_i)^2 = \frac{1}{n-2} \sum (y_i - \bar{y})^2$$

ويستج من البند (٦ - ٦) أن الإحصاءة

$$(11) \quad \frac{\beta - B}{\sigma_{\beta}} = t$$

يكون لها توزيع ت بدرجات حرية (٧ - ٧) . وبهذه الإحصاءة نستطيع كالمعتاد اختبار الفرض الصفري ف :  $\beta = k$  ضد أى فرض آخر وذلك بإيجاد ت من بيانات العينة أى بوضع  $B = k$  وعلى أساس صحة الفرض الصفري أى بوضع  $\beta = k$  ثم مقارنة هذه القيمة بالقيمة الحرجة  $t_{\alpha/2, 7}$  التى نستخرجها من جدول ت .

نتيجة : اختبار خطية العلاقة بين  $y$  ،  $x$  .

بصفة خاصة نستطيع اختبار الفرض الصفري ف :  $\beta = 0$  ضد الفرض  $\beta \neq 0$  . فإذا أشار الاختبار بقبول الفرض الصفري حكمنا بعدم وجود علاقة خطية بين  $y$  ،  $x$  ، لأن معنى  $\beta = 0$  أن تكون  $\alpha = 0$  أى  $x$  تكون ثابتة ولا يكون لقيم  $y$  أى أثر خطى على  $x$  . أما إذا رفضنا الفرض الصفري لصالح الفرض

الآخر  $\beta \neq 0$  . نحكم بأن هناك علاقة خطية بين المتغيرين وإن كان هذا لا يمنع من وجود علاقة أخرى قد تكون أفضل من العلاقة الخطية مثل  $\alpha + \beta$  وهناك اختبار آخر نعرف منه مدى انحراف العلاقة الحقيقية بين  $\alpha$  ،  $\beta$  عن العلاقة الخطية ، وسنقدم هذا الاختبار في البند (٩ - ٩) .

مثال (٩ - ٣) :

ابحث وجود علاقة خطية بين المتغيرين  $\alpha$  ،  $\beta$  من بيانات المثال (٩ - ١) مستخدماً مستوى الدلالة ٠,٠١ .

الحل :

الفرض الصفري  $F_0 : \beta = 0$  ( لا يوجد علاقة خطية )

الفرض الآخر  $F_1 : \beta \neq 0$  ( اختبار ذو جانبيين )

$$t_{\alpha} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}{n-2}}} = \frac{110,11}{\sqrt{\frac{(30,3)^2}{9}}}$$

$$= 13,1$$

$$\therefore \sqrt{t_{\alpha}^2} = 3,6194$$

كما أن  $\alpha = 0,0396$  وقد سبق إيجادها

من (١١) وعلى أساس أن  $\beta = 0$  نجد أن :

$$t_{\alpha} = \frac{3,6194 \times 2,9303}{0,0396} = 26,6002$$

من الجدول  $t_{\alpha, n-2} = 2,499$

نرفض  $F_0$  ونحكم بأن هناك علاقة خطية بين المتغيرين .

مثال (٩ - ٤) :

في بيانات المثال (٩ - ١) إبحث ما إذا كانت  $\beta < ٢,٥$  عند مستوى الدلالة ٠,٠١

الحل :

$$\text{ف : } \beta = ٢,٥$$

$$\text{ف : } \beta < ٢,٥ \quad (\text{اختبار ذو جانب واحد})$$

من (١١) وعلى أساس صحة الفرض الصفري  $\beta = ٢,٥$  نجد أن

$$٢,٨٨٦٣ = \frac{٣,٦١٩٤ \times (٢,٥ - ٢,٩٣٠٣)}{٠,٥٣٩٦} = \text{ت ي}$$

$$\text{من الجدول : ت } ٠,٠١ [٧] = ٢,٩٩٨$$

نقبل الفرض الصفري أن  $\beta = ٢,٥$  عند المستوى ٠,٠١ ونحكم بأن  $\beta$  لا تزيد عن ٢,٥ .

(ثانياً) فترات الثقة للبارامتر  $\beta$  .

من الإحصاءة (١١) نستطيع أن نثبت أن العددين

$$(١٢) \quad \text{ب} \pm \frac{\bar{y} - \mu}{\frac{s}{\sqrt{n}}} \times \alpha^{[٢ - ٥]}$$

هما حددا الثقة بدرجة  $(١ - \alpha)$  للبارامتر  $\beta$  .



(ثالثاً) فترات الثقة للقيمة الحقيقية للمتغير  $\mu$  عند قيمة معينة  $s$  :

يمكن إثبات أن العددين

$$(13) \quad \bar{x} \pm z_{\alpha/2} \sqrt{\frac{s^2}{n}}$$

هما حدا الثقة بدرجة  $(1 - \alpha)$  لقيمة  $\mu$  عندما تأخذ  $s$  القيمة  $s$ .

مثال (٩ - ٥) :

للمثال (٩ - ١) أوجد فترة ثقة بدرجة ٩٥٪ لكل من :

(أ) البارامتر  $\beta$  (ب) القيمة الحقيقية للمتغير  $\mu$  عند  $s = 2$

الحل :

(أ) نعوض في (١٢) مع ملاحظة أن  $t_{\alpha/2, n} = 2,365$

$$\text{الحدا الأدنى للفترة} = 2,365 \times \frac{0,0396}{3,6194} - 2,9303 = 2,078$$

$$\text{الحدا الأعلى للفترة} = 2,365 \times \frac{0,0396}{3,6194} + 2,9303 = 3,283$$

إذن الفترة (٢,٠٧٨ ، ٣,٢٨٣) هي فترة ثقة بدرجة ٩٥٪ للبارامتر  $\beta$ .

(ب) نعوض في (١٣) ، مع ملاحظة أنه عندما  $s = 2$  فإن

$$\hat{\mu} = 2 \times 2,9303 + 0,2068 = 6,1174$$

$$\text{الحدا الأدنى للفترة} = 2,365 \times \sqrt{\frac{(3,2667 - 2)^2}{13,1} + \frac{1}{4}} + 6,1174 = 4,689$$

$$= 4,689 = 2,365 \times 1,1197 \times 0,0396 - 6,1174$$

الحد الأعلى للفترة = ٧,٥٤٦

إذن الفترة (٤,٦٨٩ ، ٧,٥٤٦) هي فترة ثقة بدرجة ٩٥٪ لقيمة  $\mu$  عند  $n = ٢٠$

ملاحظة :

هناك برامج جاهزة للحاسب الالكتروني تعطى جميع القيم المطلوبة في تحليل الانحدار بدءاً من قيم الصيغة (٥) إلى قيم الصيغة (١٣) بمجرد تغذيته بالبيانات الخام .

(٩ - ٦) التوسع في استخدام الانحدار الخطي البسيط :

يمكن تناول بعض العلاقات غير الخطية بنفس الطريقة التي نتناول بها العلاقات الخطية وذلك باختيار تحويلات مناسبة للمتغيرات بحيث تأخذ العلاقة المعطاة الصورة الخطية . وكمثال لذلك نفرض أن العلاقة بين المتغيرين  $x$  ،  $y$  على الصورة :

$$y = \alpha + \beta x \quad \text{حيث } \alpha > 0 , \beta > 0$$

ونريد تقدير  $\alpha$  ،  $\beta$  من العينة . بأخذ لوغاريتم كل من الطرفين للأساس  $e$  نجد أن :

$$y = \alpha + \beta x \quad \text{لو } y$$

$$\bar{y} = \bar{\alpha} + \beta \bar{x} \quad \text{أو } \bar{y}$$

$$\bar{\alpha} \quad \text{حيث } \bar{\alpha} = \alpha , \quad \bar{y} = y \quad \text{لو } y$$

وهذه العلاقة الأخيرة على صورة خطية . لإيجاد تقديرين  $\alpha$  ،  $\beta$  للمجهولين ، نبدأ بتحويل كل قيمة  $s$  إلى  $l$  لو  $s$  فيصبح لدينا الأزواج (  $s$  ،  $s$  ) بدلا من الأزواج (  $s$  ،  $s$  ) . نوجد التقديرين  $\alpha$  ،  $\beta$  كالمعاد من الصيغتين ( ٥ ، ٦ ) فتكون  $\alpha$  تقديرًا للبارامتر  $\beta$  وتكون  $\beta$  هي تقدير للبارامتر  $\alpha$  وتكون معادلة الانحدار هي :

$$s = \alpha + \beta l$$

وبنفس الطريقة يمكن تناول المعادلات الآتية :

$$s = \alpha + \frac{\beta}{s} \quad \text{حيث } s \neq 0 \quad (\text{ضع } s = \frac{1}{s} \text{ التحويل القالب})$$

$$s = \alpha + \beta s \quad \text{حيث } s < 0 , \alpha < 0 , s < 0 \quad (\text{تحويل لوغاريتمي})$$

$$s = \alpha + \beta s^3 \quad \text{حيث } s = 3 \quad (\text{ضع } s = s^3)$$

$$s = \alpha + \beta s \quad \text{حيث } s < 0 , \alpha < 0 , \beta < 0 \quad (\text{تحويل لوغاريتمي})$$

مثال ( ٩ - ٦ ) :

المعروف أن العلاقة بين ضغط الغاز  $s$  وحجمه  $h$  تأخذ الصورة  $h = \alpha + \beta s$  أوجد تقديرًا للبارامتريين المجهولين  $\alpha$  ،  $\beta$  من البيانات التجريبية الآتية ثم أوجد أحسن تقدير لضغط الغاز حين يكون حجمه  $h = 100$

ح	١٩٤,٠	١١٨,٦	٨٨,٧	٧٢,٤	٦١,٨	٥٤,٣
ض	١٠,١	١٩,٢	٢٨,٤	٣٧,٦	٤٩,٥	٦١,٢

الحل :

نحول المعادلة  $\alpha = \beta$  ح إلى صورة خطية بأخذ لوغاريتم كل من الطرفين للأساس ١٠ .

$$\therefore \text{لو ض} + \beta = \text{لو ح} = \text{لو } \alpha \quad \text{أو} \quad \text{لو ض} = \text{لو } \alpha - \beta = \text{لو ح}$$

$$\text{أو} \quad \text{ص} = \bar{\alpha} + \bar{\beta} \text{ س}$$

$$\text{حيث ص} = \text{لو ض} , \bar{\alpha} = \text{لو } \alpha , \bar{\beta} = \beta - , \text{س} = \text{لو ح}$$

لتقدير بارامترات هذه المعادلة الخطية ينبغي أن نوجد لوغاريتم كل قيمة معطاة من قيم الحجم ح ولوغاريتم كل قيمة مناظرة من قيم الضغط ص . وباستخدام جداول اللوغاريتمات أو الحاسبات نحصل على العمودين الأول والثاني من الجدول الآتي ، ثم نستكمل الحساب كما في المثال (٩ - ١) .

الجدول (٩ - ٢)

س = لو ح	ص = لو ض	س'	س ص
٢,٢٨٧٨	١,٠٠٤٣	٥,٢٣٤٠	٢,٢٩٧٦
٢,٠٧٤١	١,٢٨٣٣	٤,٣٠١٩	٢,٦٦١٧
١,٩٤٧٩	١,٤٥٣٣	٣,٧٩٤٣	٢,٨٣٠٩
١,٨٥٩٧	١,٥٧٥٢	٣,٤٥٨٥	٢,٩٢٩٤
١,٧٩١٠	١,٦٩٤٦	٣,٢٠٧٧	٣,٠٣٥٠
١,٧٣٤٨	١,٧٨٦٨	٣,٠٠٩٥	٣,٠٩٩٧
١١,٦٩٥٣	٨,٧٩٧٥	٢٣,٠٠٥٩	١٦,٨٥٤٣

$$\bar{S} = 1,9492 \text{ ، } \bar{V} = 1,4663$$

على فرض أن  $\bar{A}$  ،  $\bar{B}$  هما التقديران المطلوبان للثابتين  $\alpha$  ،  $\beta$  نجد ما يلي :

$$\text{من الصيغة (5) : } \bar{B} = \frac{8,7970 \times 11,7903 - 16,8043 \times 6}{(11,7903) - 23,009 \times 6}$$

$$= 1,40 =$$

$$\text{من الصيغة (6) : } \bar{A} = \bar{S} - \bar{B} \bar{V} = 1,4663 - 1,40 \times 1,9492$$

$$= 4,1902 = 4,2 \text{ تقريباً}$$

$$\text{إذن } \bar{B} = \bar{V} = 1,40$$

$$\bar{A} = 10,10 = 10,1489 = 10,1489 \text{ تقريباً}$$

وهذان هما التقديرات المطلوبان للثابتين  $\alpha$  ،  $\beta$  وعلى ذلك فإن المعادلة التي تربط الحجم والضغط الناتجة من العينة هي :

$$V^{10,10} = 10,1489$$

وحين  $H = 100$  فإن أحسن تقدير للضغط ض يتبع كالاتي :

$$V^{10,10} \times 100 = 10,1489 \text{ ومنها ض } \times 10,1489 = 28,10$$

$$\text{إذن ض } = 25,12 \text{ تقريباً .}$$

## تمارين (٩ - ١)

من كل من البيانات المبينة في المسائل الخمسة الآتية :

- (أ) أوجد معادلة انحدار  $\bar{y}$  على  $\bar{x}$  .  
 (ب) أوجد الخطأ المعياري  $\bar{e}$  لخط الانحدار .  
 (ج) ابحث ما إذا كان هناك علاقة خطية بين المتغيرين  $\bar{x}$  ،  $\bar{y}$  .  
 (د) إذا ثبت وجود علاقة خطية فأوجد فترة ثقة بدرجة ٩٥٪ للقيمة الحقيقية للمتغير  $\bar{y}$  عند القيمة  $\bar{x}$  المعطاة .

٣,٤	٣,٣	٣,٢	٣,٢	٣,٢	٣,١	٣,٠	٢,٩	٢,٨	س :
٣٠	٣٣	٣٤	٣٢	٣٠	٢٨	٣٠	٢٣	٢٧	ص :

حيث  $\bar{x}$  ترمز إلى كثافة الحديد الخام ( جرام /  $\text{سم}^3$  ) ،  
 $\bar{y}$  ترمز إلى النسبة المئوية لمحتوى الحديد .  
 $\bar{e} = ٣$  ،

٥٢	٥١	٥٠	٤٩	٤٨	س :
٢٨٤,٨	٢٨٣,٢	٢٨١,٤	٢٧٩,٣	٢٧٧,١	ص :

حيث  $\bar{x}$  ترمز إلى طول الطفل عند الولادة بالسنتيمترات ،  
 $\bar{y}$  ترمز إلى مدة الحمل بالأيام .  
 $\bar{e} = ٥٠$  ،

١٦	١٦	١٦	١٤	١٤	١٣	١٢	٩	٩	٨	س :
٢٤-	٢٧-	٣٥-	٣٧-	٣٥-	٥٠-	٥٠-	٥٣-	٥٥-	٥٥-	ص :

حيث  $s$  ترمز إلى درجات الحرارة بالاستيجراد  
 ،  $s$  ترمز إلى الانحراف ( بمضاعفات  $\frac{1}{1000}$  زاوية دائرية ) لنوع معين  
 من الرؤية التلسكوبية .  
 $s = 13,5$  ،

(٤)  $s$  : ١٥ ٣٠ ٦٠ ١٢٠ ٣٠٠ ٦٠٠  
 $s$  : ٤,٠ ٦,٣ ٥,٢ ٦,٤ ٥,٥ ٦,٤

حيث  $s$  ترمز إلى المدة بالثواني  
 ،  $s$  ترمز إلى مقدار المادة التي تهددت ( بالجرامات في اللتر ) من استحلاب  
 الزيت في محلول زيت الزيتون بمتأثير الاهتزاز الناتج من الصوت عن ذهبات  
 فوق الصوتية في المدة  $s$  .  
 $s = 70$  ،

(٥)  $s$  : ٠,١ ٠,٢ ٠,٤ ٠,٠٨ ٠,١٥ ٠,٣٠ ٠,٥٠ ١,٠٠ ٢,٠٠  
 $s$  : ١٢ ١٨ ٣ ٢٧ ١٦ ١٣ ١٠ ٧٠ ٩٠

حيث  $s$  ترمز إلى النسبة المئوية للدرجة تركيز الكلورونافثالين  
 ،  $s$  ترمز إلى النسبة المئوية لموت الحمل الأبيض .  
 $s = 0,25$  ،

(٦) في عينة عشوائية من ٨ أزواج من القيم ( $s_1$  ،  $s_2$ ) وجد أن معامل الانحدار  
 ب = ٠,٣١ وأن  $s_1 = ٠,٤$  ،  $s_2 = ٠,٣$

اختبر الفرض أن  $\beta = 0$  ضد الفرض أن  $\beta < 0$  .

(٩ - ٧) معنى آخر للانحدار - تحليل التباين :

في البنود السابقة من هذا الفصل كنا نأخذ الانحدار على أنه وسيلة لإيجاد معادلة تمكنا من معرفة أحسن تقدير لقيمة متغير عشوائى  $y$  عن طريق قيمة معطاة لمتغير  $x$  . ولذلك وصفنا معادلة الانحدار بأنها معادلة تنبؤ . إلا أن الانحدار يؤخذ أيضا على أنه وسيلة لتفسير الاختلاف المشاهد في قيم المتغير  $y$  ، وذلك استجابة لتساؤل هام عن العوامل المؤثرة في هذا الاختلاف ، وبصفة خاصة عما إذا كان هذا الاختلاف يرجع إلى التغير في قيم المتغير  $x$  أم يرجع إلى عوامل عشوائية أو عوامل أخرى لا تتعلق بالمتغير  $x$  .

إن الإجابة عن هذا التساؤل تتطلب تحليل الاختلاف في قيم  $y$  وهو  $\sum (y_i - \bar{y})^2$  =  $\sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$  إلى مركبتين مستقلتين تعتمد أحدهما على قيم المتغير  $x$  ولا تعتمد الأخرى عليها ثم تقييم كل من هاتين المركبتين .

وليبيان كيفية هذا التحليل نعتبر المساواة الآتية :

$$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 \quad \text{حيث } i = 1, 2, \dots, n$$

إن هذه المساواة واضحة جبريا ، كما أنها تتضح هندسيا من الشكل (٩ - ٥) مع ملاحظة أن النقطة  $(\bar{x}, \bar{y})$  تقع دائما على خط الانحدار . وبترتيب طرفي المساواة ثم الجمع نتج المساواة الآتية :

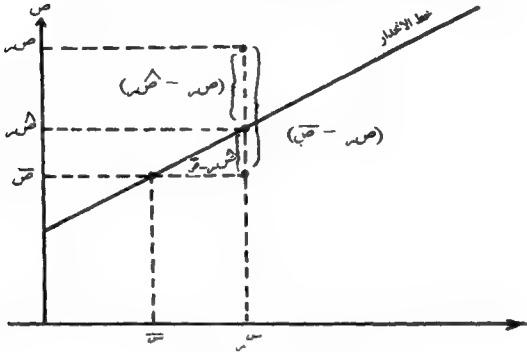
$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (١٤)$$

مع ملاحظة أن الحد الأوسط في عملية تجميع الطرف الأيسر يتعلم عند عملية



الجمع . وبذلك نكون قد جزأنا الاختلاف الكلي في ص إلى مركبتين بطريقة مشابهة لتقسيم الاختلاف الكلي في تحليل التباين .

لنبحث الآن في المعنى الذي تتضمنه كل من هاتين المركبتين .



الشكل (٩ - ٥) : تجزء الاختلاف في الطير ص

### الاختلاف المفسر والاختلاف غير المفسر

#### EXPLAINED AND UNEXPLAINED VARIATION

من الشكل (٩ - ٥) نرى أننا إذا رسمنا خط الانحدار وحددنا قيمة معينة  $\bar{s}$  فإن قيمة  $\bar{v}$  المناظرة تكون قد تحددت تماماً لأنها تقع على خط الانحدار ويكون الفرق  $(\bar{v} - \hat{v})$  بين  $\bar{v}$  والقيمة الثابتة  $\bar{v}$  هو فرق يعتمد كلية على قيمة  $\bar{s}$  وبالتالي فإن مجموع مربعات الفروق  $\sum (\bar{v} - \hat{v})^2$  يعتمد كلية على قيم المتغير  $s$  . ولما كان هذا المقدار هو جزء من الاختلاف الكلي في ص كما يظهر من المساواة (١٤) فإن  $\sum (\bar{v} - \hat{v})^2$  يكون هو الجزء من الاختلاف في ص

الذى يُعزى إلى التغير في  $s$  ، أى إلى التغير الذى حدث في  $v$  نتيجة للتغير في  $s$  ، وهذا ما تسمية بانحدار  $v$  على  $s$  ولذلك نرمز لهذا الاختلاف بالرمز  $m$  ( الانحدار الخطى ) وله درجة واحدة من درجات الحرية . ولما كان هذا الاختلاف مصدره معروف ( وهو التغير في  $s$  ) فقد اصطلح على تسميته أيضا بالاختلاف المفسر ونكتبه رمزيا كالآتي :

الاختلاف المفسر =  $m$  (الانحدار) =  $\bar{v} - \bar{v}_s$  (  $\bar{v}_s$  -  $\bar{v}$  ) بدرجة حرية واحدة (١٥) ويمكن إثبات أن هذا الاختلاف يمكن أن يحسب كالآتي :

الاختلاف المفسر =  $\bar{v} - \bar{v}_s$  حيث  $\bar{v}_s$  معامل الانحدار (١٦)

$$\text{أو} \quad \frac{[\sum (v - \bar{v}_s)^2]}{\sum (v - \bar{v}_s)^2} = \quad (١٧)$$

$$\text{حيث } \bar{v}_s = \frac{\sum (v - \bar{v}_s)}{n} = \bar{v} - \bar{v}_s$$

$$\bar{v} - \bar{v}_s = \frac{\sum (v - \bar{v}_s)^2}{\sum (v - \bar{v}_s)^2} = \bar{v} - \bar{v}_s$$

أما المركبة الثانية وهى  $\bar{v} - \bar{v}_s$  فهى الاختلاف الذى يتبقى بعد طرح الاختلاف المفسر من الاختلاف في  $v$  ، وهو يعتمد على عوامل مجهولة لا تتعلق بالتغير  $s$  . وقد اصطلح على تسمية هذا الاختلاف بالاختلاف المتبقى residual variation أو بالاختلاف غير المفسر . ومع تذكر أن  $\bar{v}_s$  هى قيم عشوائية فإن هذا الاختلاف يعبر عن التشتت غير المنتظم لنقط شكل الانتشار حول خط الانحدار أى عن الانحرافات الرأسية حول خط الانحدار وسنرمز له بالرمز  $m$  ( الانحراف عن خط الانحدار ) وله  $n - 2$  من درجات الحرية . أى أن

الاختلاف غير المفسر = ٢٢ ( الانحراف عن خط الانحدار )

$$= \sum (\text{صر} - \hat{\text{صر}})^2$$

$$= ٢٢ (\text{صر}) - \text{الاختلاف المفسر} \quad (١٨)$$

والمعنى الذى يتضمنه هذا الاختلاف يرر استخدامه كأساس لحساب الخطأ  
المعيارى  $\sigma_{\text{صر}}$  الوارد فى الصيغة (٩) بالبند (٩ - ٤) ، حيث

$$(١٩) \quad \sigma_{\text{صر}} = \frac{\sum (\text{صر} - \hat{\text{صر}})^2}{\text{الاختلاف غير المفسر}} = \frac{\sum (\text{صر} - \hat{\text{صر}})^2}{\text{درجات الحرية}}$$

نستطيع حيثذ أن نكتب المتساوية (١٤) كالآتى :

$$٢٢ (\text{صر}) = ٢٢ (\text{الانحدار الخطى}) + ٢٢ (\text{الانحراف عن خط الانحدار}) \quad (٢٠)$$

بدرجات حرية  $١ - ١$  ،  $١ - ١$  ،  $٢ - ١$  على الترتيب .

ويفضل تسجيل قيم هذه المتساوية فى جدول التباين الآتى .

الجدول (٩ - ٣)

جدول التباين للانحدار

مصدر التباين	٢٢	د ح	ط ٢	فى
الاختلاف المفسر (الانحدار الخطى)	$\sum (\hat{\text{صر}} - \text{صر})^2$	١	$\frac{١}{٢٢} \sum \text{صر}^2$	$\frac{١}{٢٢} \sum \text{صر}^2$
الاختلاف غير المفسر (الانحراف عن الخطية)	$\sum (\text{صر} - \hat{\text{صر}})^2$	$٢ - ١$	$\frac{١}{٢٢} \sum \text{صر}^2$	$\frac{١}{٢٢} \sum \text{صر}^2$
الاختلاف الكلى فى ص	$\sum (\text{صر} - \bar{\text{صر}})^2$	$١ - ١$		

بعد تجزئ الاختلاف الكلى فى ص بهذه الصورة يبقى أن نختبر ما إذا كان الانحدار الخطى قد فسر جزءا ذا بال من هذا الاختلاف ، أى أن نختبر ما إذا كان التباين المفسر أكبر كبرا جوهريا من تباين الاختلاف غير المفسر . وهذا ما نقيسه باختبار ف بالصيغة الآتية بشرط توفر افتراضات الانحدار ، ومع ملاحظة أن هذين التباينين هما تقديران مستقلان للتباين  $\sigma^2$  لتوزيع المتغير  $v$  .

$$F = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} \quad \text{بلرجتى حرية } ١، n - ٢ \quad (٢١)$$

وهذا الاختبار يكافئ اختبار ت الوارد بالصيغة (١١) بالبند (٩ - ٥) لاختبار وجود علاقة خطية بين المتغيرين  $v$  ،  $w$  أى لاختبار الفرض الصفري  $\beta = ٠$  ضد الفرض  $\beta \neq ٠$  . ويمكن اشتقاق أى منهما من الآخر ، إذ أن قيمة ف بالصيغة (٢١) تساوى مربع قيمة ت بالصيغة (١١) . وبذلك نكون قد توصلنا إلى طريقة أخرى لاختبار وجود علاقة خطية بين المتغيرين .

وجدير بالذكر أن الصيغة (٢١) هى صيغة عامة لاختبار مدى دقة التنبؤ من خط الانحدار مهما كان عدد المتغيرات المستقلة المستخدمة فى عملية التنبؤ، وتختلف طريقة حساب البسط والمقام فى هذه الصيغة باختلاف عدد المتغيرات التنبؤية .

مثال (٩ - ٧) :

ابحث وجود علاقة خطية بين المتغيرين  $v$  ،  $w$  من بيانات المثال (٩ - ١) مستخدما طريقة تحليل التباين .

الحل :

الفرض الصفري :  $\beta = ٠$       الفرض الآخر :  $\beta \neq ٠$

من الجدول (٩ - ١) نستطيع حساب القيم اللازمة لاستخدام الصيغة (٢١) كالآتي :

$$٢٢ (ص) = \frac{٩١,١}{٩} - ١٠٣٦,٦٥ = ١١٤,٥١٥٦$$

$$٢٢ (س) = \frac{٣٠,٣}{٩} - ١١٥,١١ = ١٣,١$$

$$٢ ص (س) = \frac{٩١,١ \times ٣٠,٣}{٩} - ٣٤٥,٠٩ = ٣٨,٣٨٦٧$$

$$\text{من (١٧) : الاختلاف المفسر} = \frac{٣٨,٣٨٦٧}{١٣,١} = ٢,٩٤٨٣٩ = ٢,٩٤٨٣٩$$

$$\therefore \text{الاختلاف غير المفسر} = ١١٤,٥١٥٦ - ٢,٩٤٨٣٩ = ١١٢,٥٦٧٢$$

$$٧ = ٢,٩٤٨٣٩$$

$$٢,٠٣١٧ =$$

وينشأ جدول التباين الآتي :

الجدول (٩ - ٤)

مصدر التباين	٢٢	د ح	ط ٢	في
الاختلاف المفسر	١١٢,٤٨٣٩	١	١١٢,٤٨٣٩	٣٨٧,٦٠٨
الاختلاف غير المفسر	٢,٠٣١٧	٧	٠,٢٩٠٧	
الاختلاف الكلي	١١٤,٥١٥٦	٨		

إن القيمة  $F = 387,608$  أكبر بكثير من القيمة الحرجة  $F_{0.01} = 12,2$  مما يجعلنا نرفض الفرض الصفري عند مستوى عالى من الدلالة ونحكم بوجود علاقة خطية بين المتغيرين .

نلاحظ أن الجذر التربيعى لقيمة  $F$  وهو  $\sqrt{387,608} = 19,688$  يساوى قيمة  $t = 19,652$  السابق إيجادها بالمثال (٩ - ٣) ويرجع الفرق بين القيمتين إلى أخطاء التقريب .

### معامل التحديد COEFFICIENT OF DETERMINATION

استكمالا للانحدار كوسيلة لتفسير الاختلاف في قيم المتغير التابع  $y$  يهنا أن نقدر نسبة الاختلاف الذى فسره الانحدار الخطى إلى الاختلاف الكلى فى المتغير  $y$  . وهذه النسبة تسمى بمعامل التحديد ويرمز لها بالرمز  $r^2$  . أى أن :

$$r^2 = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} \quad (٢٢)$$

فى المثال السابق نجد أن :

$$r^2 = \frac{112,4839}{114,5156} = 0,982 \text{ تقريبا}$$

وهذا يعنى أن الانحدار الخطى قد فسر حوالى ٩٨,٢% من الاختلاف الكلى أى جزءا جوهريا منه ، مما يؤكد صحة العلاقة الخطية بين المتغيرين . ويرمز لمعامل التحديد بالرمز  $r^2$  لأنه يساوى مربع معامل الارتباط  $r$  الذى سنتناوله فى الفصل التالى . ويلاحظ أن

$$0 \leq r^2 \leq 1$$

لأن كلا من البسط والمقام فى التعريف (٢٢) هو مجموع مربعات لا يمكن أن يكون سالبا وإذن  $0 \leq r^2$  كما أن البسط جزء من المقام كما يتضح من المتساوية

(١٤) واذا  $r \geq ١$  . وإذا كانت  $r = ٠$  فإن هذا يعنى أن الانحدار الخطى لا يفسر شيئاً من الاختلاف فى ص ولا يكون للمتغير فى س أى أثر فى التغير فى ص . أما إذا كانت  $r = ١$  فإن هذا يعنى أن الانحدار قد فسر الاختلاف فى ص بأكمله ، وهذا يحدث حين تكون القيم المشاهدة ص مساوية للقيم المناظرة ص المقدرة من معادلة الانحدار ، أى تكون النقط (س ، ص) المشاهدة واقعة جميعها على خط الانحدار . وحين تكون قيمة  $r$  صغيرة فإن هذا يعنى أن الجزء الأكبر من الاختلاف فى ص يرجع إلى عوامل ومتغيرات لا تدخل فى الانحدار أى لا علاقة لها بالتغير فى المتغير المستقل ص .

من الصيغة (١٧) يمكن أن نكتب معامل التحديد بدلالة مجاميع المربعات ومجاميع حواصل الضرب كالتالى :

$$(٢٣) \quad \frac{r^2 [(ص ، س) - (ص)(س)]}{(ص) ٢٢ . (س) ٢٢} = r^2$$

هذا ويمكن كتابة اختبار ف بالصيغة (٢١) بدلالة معامل التحديد كالتالى :

$$F = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} = \frac{\text{الاختلاف المفسر} \div ١}{\text{الاختلاف غير المفسر} \div (٢ - ١)}$$

$$= \frac{\text{الاختلاف المفسر}}{(\text{الاختلاف الكلى} - \text{الاختلاف المفسر}) \div (٢ - ١)}$$

وبقسمة كل من البسط والمقام على الاختلاف الكلى  $r^2 (ص)$  ينتج أن

$$(٢٤) \quad F = \frac{r^2}{(١ - r^2) / (٢ - ١)}$$

وهذه صيغة ثالثة لاختبار وجود علاقة خطية بين متغيرين .

(٩ - ٨) تحليل الانحدار حين يكون هناك أكثر من قيمة ص لكل قيمة س .

في الأمثلة والتمارين السابقة كانت التجربة تحدد قيمة عشوائية واحدة ص لكل قيمة ثابتة س . إلا أن بعض الأبحاث تفضل تصميم التجربة بحيث نحصل منها على أكثر من قيمة عشوائية من المتغير ص لكل قيمة من قيم المتغير ص ، خاصة وأن هذا التصميم يوفر لنا الفرصة للحكم على جودة العلاقة المفروضة بين المتغيرين كما سنرى في البند (٩ - ٩) .

فإذا كان هناك ك من القيم السينية  $S_1, S_2, \dots, S_p$  ، يكون لدينا ك من المجموعات الصادية المناظرة لها ، وتتخذ البيانات في هذه الحالة الصورة المبينة بالجدول (٩ - ٥) والتي تشبه الصورة التي يتخذها تحليل التباين للتجارب ذوات العامل الواحد .

الجدول (٩ - ٥)

$S_1$	$S_2$	...	$S_p$	...	$S_q$
ص <sub>١١</sub>	ص <sub>٢١</sub>	...	ص <sub>١١</sub>	...	ص <sub>١١</sub>
ص <sub>١٢</sub>	ص <sub>٢٢</sub>	...	ص <sub>١٢</sub>	...	ص <sub>١٢</sub>
...	...	...	...	...	...
...	...	...	...	...	...
ص <sub>١٤</sub>	ص <sub>٢٤</sub>	...	ص <sub>١٤</sub>	...	ص <sub>١٤</sub>
...	...	...	...	...	...
...	...	...	...	...	...
ص <sub>١٥</sub>	ص <sub>٢٥</sub>	...	ص <sub>١٥</sub>	...	ص <sub>١٥</sub>

وإذا تبيننا نفس الافتراضات الثلاثة للانحدار الخطي فإن أسلوب التحليل يسم على نفس النمط السابق تقديمه مع بعض التعديلات التي يقتضيها الوضع الجديد للبيانات كما يتبين من المثال الآتي .



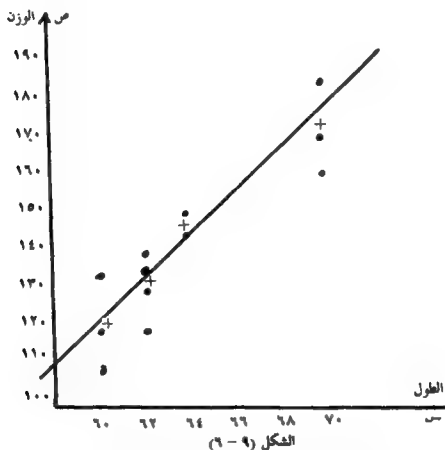
مثال (٩ - ٨) :

في دراسة عن توزيع الأوزان لمجتمع من الرجال وعلاقة هذا التوزيع بالأطوال ، قسم مجتمع الرجال من حيث الطول إلى ٤ أقسام تتساوى فيها الأطوال بالتقريب وتمثلها الأطوال ٦٠ ، ٦٢ ، ٦٤ ، ٧٠ . وفي كل من هذه الأطوال لاختير عدد من الرجال عشوائيا وقيست أوزانهم وسجلت المقاييس في الجزء العلوى من الجدول (٩ - ٦) الآتى :

الجدول (٩ - ٦)

الأطوال					الأوزان
س <sup>١</sup>	س <sup>٢</sup>	س <sup>٣</sup>	س <sup>٤</sup>		
٦٠	٦٤	٦٢	٧٠		
١١٠	١٢٠	١٥٠	١٧٠		
١٣٥	١٤٠	١٤٥	١٨٥		
١٢٠	١٣٠		١٦٠		
	١٣٥				
ن <sup>١</sup>	ن <sup>٢</sup>	ن <sup>٣</sup>	ن <sup>٤</sup>	ن <sup>٥</sup>	
٣٦٥	٥٢٥	٢٩٥	٥١٥	١٢ = ن	مجموع ص <sup>١</sup>
٤٤٧٢٥	٦٩١٢٥	٤٣٥٢٥	٨٨٧٢٥	٢٤٦١٠٠ = ن	مجموع ص <sup>٢</sup>
١٨٠	٢٤٨	١٢٨	٢١٠	٧٦٦ = ن	مجموع ص <sup>٣</sup>
١٠٨٠٠	١٥٣٧٦	٨١٩٢	١٤٧٠٠	٤٩٠٦٨ = ن	مجموع ص <sup>٤</sup>
٢١٩٠٠	٣٢٥٥٠	١٨٨٨٠	٣٦٠٥٠	١٠٩٢٨٠ = ن	مجموع ص <sup>٥</sup>

نعتبر عن هذه البيانات هندسيا كما في الشكل (٩ - ٦) الذي يعرض شكل الانتشار ويشتمل على ١٢ نقطة تشترك بعضها في الإحداثيات السينية وكل نقطة تعبر عن طول ووزن أحد الرجال . أما النقط المشار إليها بالعلامة + فتمثل متوسطات المجموعات الصادية عند القيم السينية المناظرة أى تمثل النقط (س<sub>١</sub> ، ص<sub>١</sub>) ، (س<sub>٢</sub> ، ص<sub>٢</sub>) ، (س<sub>٣</sub> ، ص<sub>٣</sub>) ، (س<sub>٤</sub> ، ص<sub>٤</sub>) .



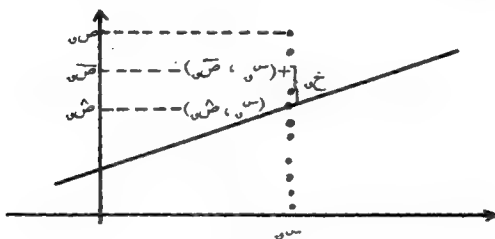
شكل الانتشار وسط الانحدار لبيانات المثال (٩ - ٨)

المطلوب في هذا المثال إيجاد معادلة انحدار ص على س واختبار دلالة هذا الانحدار .

**إيجاد معادلة الانحدار :**

في المثال (٩ - ٧) حيث كان لدينا قيمة واحدة ص لكل قيمة س كان بحثنا يهدف إلى معرفة ما إذا كانت القيم الصادية ص<sub>١</sub> ، ص<sub>٢</sub> ، ... تقع على خط مستقيم هو خط الانحدار،

أما في المثال (٩ - ٨) حيث لدينا مجموعة من القيم الصادية لكل قيمة  $s$  فإن بحثنا يهدف في هذه الحالة إلى معرفة ما إذا كانت المتوسطات  $\bar{v}_s$  ،  $\bar{v}_{s+1}$  ، ... للمجموعات الصادية تقع على خط مستقيم . وعلى ذلك فإنه حسب مبدأ المربعات الصغرى - راجع البند (٩ - ٣) - يكون خط الانحدار هو ذلك الذى نختار ثوابته بحيث تجعل الدالة  $\chi^2 = \sum (v_s - \bar{v}_s)^2$  نهاية صغرى ، حيث  $\bar{v}_s$  هو الإحداثى الصادى للنقطة الواقعة على خط الانحدار والتى إحداثيها السينى  $s$  - انظر الشكل (٩ - ٧) .



الشكل (٩ - ٧)

وبهنا أن نشير إلى أننا إذا وضعنا بدلا من كل قيمة  $v_s$  فى أى عمود  $v$  الوسط الحسابى  $\bar{v}_s$  للصادات فى هذا العمود فإن معادلة الانحدار التى تنتج تكون هى بناتها معادلة انحدار  $v$  على  $s$  لأن هذا التغير لا يؤثر فى الجاميع أو مجاميع المربعات أو مجاميع حواصل الضرب ، وهذا ما يمكن بيانه جبريا . وهذه الحقيقة تعنى أنه إذا وقعت متوسطات الأعمدة على خط مستقيم فإن هذا الخط ينطبق على خط انحدار  $v$  على  $s$  .

لايجاد معامل الانحدار  $b$  نستخدم الصيغة (٥) التى تأخذ فى هذه الحالة الشكل الآتى :

$$\frac{\frac{(س س س س)(س س س س)}{٧} - س س س س}{\frac{(س س س س)}{٧} - س س س س} = س$$

وفي هذا المثال نجد أن :

$$٥,٠٢٩٢ = \frac{٨٦٣, ٣٤}{١٧١,٦٦٧} = \frac{\frac{١٧٠٠ \times ٧٦٦}{١٢} - ١٠٩٣٨٠}{\frac{٧٦٦}{١٢} - ٤٩٠٦٨} = س$$

$$٦٣,٨٣٣٣ = \frac{٧٦٦}{١٢} = س س س س = \frac{١}{٧}$$

$$١٤١,٦٦٦٧ = \frac{١٧٠٠}{١٢} = س س س س = \frac{١}{٧}$$

∴ معادلة الانحدار س على س مقربة إلى ٣ خانات عشرية هي :

$$س = ١٤١,٦٦٧ + ٥,٠٢٩ (س - ٦٣,٨٣٣)$$

$$س = ١٧٩,٣٤٩ - ٥,٠٢٩ س$$

اختبار وجود علاقة خطية :

على فرض توفر شروط الانحدار يمكننا اختبار وجود علاقة خطية بين المتغيرين  
أى اختبار الفرض الصفري  $\beta = ٠$  ضد الفرض  $\beta \neq ٠$  باستخدام اختبار ت  
بالصيغة (١١) أو اختبار ف بالصيغة العامة (٢١) أو بالصيغة المكافئة (٢٤) .  
ونستعمل هنا الصيغة العامة وهى :

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} = \text{بلر جتى حرية ١ ، ٧ - ٢}$$

من بيانات المثال نجد مايلي :

$$\frac{{}^1(٤٤ ص) - {}^1(٤٤ ص)}{٥} = ٢٢ (ص) = ٢٤٦١٠٠ - \frac{{}^1١٧٠٠}{١٢} = ٥٢٦٦,٦٧$$

$$\frac{{}^2(٢ ص - ٣ ص ، ص)}{{}^2(٢ ص)} = \text{الاختلاف المفسر}$$

$$٤٣٤١,٨٧٠٨ = \frac{{}^1(٨٦٣,٣٤)}{١٧١,٦٦٧} = \frac{[\frac{{}^1١٧٠٠ \times ٧٦٦}{١٢} - ١٠٩٣٨.]}{\frac{{}^1٧٦٦}{١٢} - ٤٩٠٦٨} =$$

∴ الاختلاف غير المفسر = ٤٣٤١,٨٧٠٨ - ٥٢٦٦,٦٧ = ٩٢٤,٧٩٩٢  
وبذلك نحصل على جدول التباين الآتي :

الجدول (٩-٧)

مصدر التباين	٢٢	د ح	ط ٢	في
الاختلاف المفسر	٤٣٤١,٨٧٠٨	١	٤٣٤١,٨٧٠٨	٤٦,٩٥ <sup>٥٥</sup>
الاختلاف غير المفسر	٩٢٤,٧٩٩٢	١٠	٩٢,٤٧٩٩	
الاختلاف الكلي	٥٢٦٦,٦٧٠٠	١١		

بما أن ٤٦,٩٥ أكبر بكثير من القيمة الحرجة ف  $_{[١٠,٠١٧,٠٠١]} ١٠,٠٠ =$   
نرفض الفرض الصفري  $\beta = ٠$  عند مستوى عالى من الدلالة ونحكم بوجود علاقة  
خطية بين طول الرجل ووزنه .

#### (٩ - ٩) اختبار جودة العلاقة الخطية :

نعلم أننا إذا قبلنا الفرض الصفري  $\beta = 0$  فإننا نحكم بعدم وجود علاقة خطية بين المتغيرين ولا نحتاج حينئذ إلى إجراء أى اختبارات أخرى تتعلق بخطية هذه العلاقة . أما إذا رفضنا هذا الفرض فإننا نحكم بوجود علاقة خطية بين المتغيرين لأن الانحدار الخطي يكون قد فسر جزءا ذا دلالة من الاختلاف الكلي . غير أن هذا لا يعنى أن العلاقة الخطية هي أحسن علاقة تصف العلاقة الحقيقية بين المتغيرين ، فقد تكون هناك علاقة أخرى مثل  $\alpha + \beta_1 x + \beta_2 x^2$  تفضل العلاقة الخطية في ذلك . ويستلزم الأمر هنا الاستمرار في تحليل البيانات لتقدير درجة جودة العلاقة الخطية ، وذلك بتقييم الاختلاف غير المفسر الذى يعبر عن الانحراف عن الخطية ، فإذا كان هذا الاختلاف غير ذي دلالة أى يرجع إلى العوامل العشوائية فإن العلاقة الخطية تكون هي أحسن العلاقات ، أما إذا كان هذا الاختلاف ذا دلالة فإن العلاقة الخطية لا تكون هي العلاقة المثل بين المتغيرين .

ولكن كيف نقيم الاختلاف المعبر عن الانحراف عن الخطية ؟ إننا نحتاج هنا إلى مقياس نقيس به دلالة هذا الانحراف . ولتحقيق هذا الفرض ينبنى تصميم تجربة نحصل منها على بيانات بحيث يناظر كل قيمة من القيم السينية مجموعة من القيم الصادية كما في المثال (٩ - ٨) السابق ، لأن وجود هذه المجموعات الصادية يتيح لنا فرصة إيجاد الاختلاف داخل هذه المجموعات وهو الذى يعبر عن خطأ التجريب ، وبالتالي يمكن اتخاذه معيارا لمدى دلالة الانحراف عن الخطية .

وعلى ذلك ، وعلى فرض أن لدينا بيانات من مثل هذه التجربة نبدأ بخطوة هامة هي تحليل التباين للمجموعات الصادية ، أى فصل الاختلاف الكلي في القيم الصادية ( كالمعاد ) إلى مركبتين مستقلتين تعبر الأولى عن الاختلاف بين المجموعات وتعبر الثانية عن الاختلاف داخل المجموعات ( خطأ التجريب ) ، وهذه الخطوة تتخذ الشكل المعاد الآتي :

٢٢ (ص) = ٢٢ (بين المجموعات) + ٢٢ (داخل المجموعات)

بدرجات حرية ١ - ك ، ١ - ك ، ١ - ك على الترتيب .

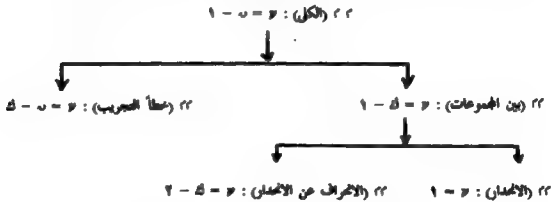
وجدير بالإشارة هنا إلى أن المتغير المستقل - في الانحدار هو متغير كمي بينما المتغير المستقل (عامل التجريب) في تحليل التباين هو متغير نوعي . وحين نجري تحليل التباين للانحدار نعتبر أن القيم العددية للمتغير - هي مجرد إشارات تعبر عن مستويات عامل التجريب .

ولما كان الهدف من تحليل الانحدار اختبار ما إذا كانت متوسطات المجموعات الصادية تقع على خط مستقيم فإن الاختلاف الذي ينبغي تحليله هو الاختلاف بين المجموعات . وعلى ذلك فإن الخطوة الثانية هي تحليل الاختلاف بين المجموعات إلى مركبتين مستقلتين تعبر الأولى عن الاختلاف المفسر : ٢٢ (الانحدار الخطي) وتعبر الثانية عن الاختلاف غير المفسر : ٢٢ (الانحراف عن الانحدار الخطي) وهذه الخطوة تتخذ الشكل الآتي :

٢٢ (بين المجموعات) = ٢٢ (الانحدار الخطي) + ٢٢ (الانحراف عن خط الانحدار)

بدرجات حرية ١ - ك ، ١ ، ١ - ك على الترتيب .

الشكل (٩ - ٨) الآتي يلخص خطوات التحليل السابق ذكرهما .



الشكل (٩ - ٨) : خطوات اختبار جودة الانحدار الخطي

كما أن الجدول (٩ - ٨) الآتي هو جدول التباين لعملية التحليل .

الجدول (٩ - ٨)

درجات الحرية	مجموع المربعات	مصدر التباين
ك - ١ ١ ك - ٢ ١ - ك	$\left. \begin{aligned} & \sum (u_i - \bar{u})^2 \\ & \sum (v_i - \bar{v})^2 \\ & \sum (w_i - \bar{w})^2 \end{aligned} \right\}$	بين المجموعات الاختلاف الحظي الاختلاف عن خط الانحدار خطأ التجريب (داخل المجموعات)
١ - ١	$\sum (u_i - \bar{u})^2$	الكل

من هذا التحليل نستطيع أن نختبر ثلاثة أمور هي :

(أولاً) اختبار دلالة الاختلاف المشاهد بين المجموعات :

أى اختبار ما إذا كانت متوسطات المجموعات الصادية تختلف باختلاف القيم السنية . والاختبار الذى يصلح لذلك هو اختبار ف المعتمد حيث

$$F = \frac{\text{التباين بين المجموعات}}{\text{التباين داخل المجموعات}} \text{ بدرجتي حرية ك - ١ ، ١ - ك (٢٥)}$$

والفرض الصفري هنا هو أن المتوسطات متساوية جميعها . فإذا قلنا هذا الفرض نحكم بعدم وجود أى علاقة بين التغير في  $\bar{u}$  والتغير في  $\bar{v}$  ويتوقف البحث عند هذا الحد . أما إذا رفضنا الفرض الصفري فنحكم بأن التغير في  $\bar{u}$  يؤثر في التغير في  $\bar{v}$  وينبغي حينئذ أن نستمر في البحث بحسب الخطوة الثانية .



(ثانياً) بحث العلاقة الخطية بين المتغيرين :

(أ) اختبار وجود علاقة خطية :

أى اختبار الفرض الصفري  $\beta = 0$  ضد الفرض  $\beta \neq 0$  وكما فعلنا في المثال (٩ - ٨) نستخدم الصيغة (٢١) وهى :

$$F = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}}$$

مع ملاحظة أن الاختلاف غير المفسر هو ذلك الاختلاف الذى لا يعتمد على المتغير  $x$  وهو يشمل الاختلاف الناشئ عن الانحراف عن خط الانحدار كما يشمل الاختلاف الناشئ عن خطأ التجريب ، وعلى ذلك فإن التعويض في هذه الصيغة من بيانات الجدول (٩ - ٨) يكون كالتالى :

$$F = \frac{٢٢ (\text{الانحدار الخطى}) \div ١}{[٢٢ (\text{الانحراف عن الخطية}) + ٢٢ (\text{خطأ التجريب})] \div (٢-١)} \quad (٢٦)$$

بدرجتي حرية ١ ،  $\alpha = ٠.٠٥$  .

وكما سبق القول في (أولاً) ليس هناك ما يدعو للقيام بهذا الاختبار أو بالاختبار الذى سيؤدي (ب) إلا إذا ثبت من الخطوة السابقة دلالة الاختلاف بين المجموعات الصادية .

(ب) اختبار جودة العلاقة الخطية :

أى اختبار دلالة الانحراف عن خط الانحدار مقدراً بمجموع مربعات الانحرافات  $\sum (y_i - \hat{y}_i)^2$  ، أو بمعنى آخر اختبار ما إذا كانت متوسطات المجموعات الصادية تقع على خط الانحدار أو قريبة منه أم تنتشر بعيدة عنه أبعاداً جوهرية . والاختبار الذى يصلح لذلك هو اختبار  $F$  بالصورة

$$F = \frac{\text{تباين الانحراف عن الخطية}}{\text{تباين خطأ التجريب}} \quad (27)$$

بدرجتي حرية ك - ٢ ، هـ - ك .

في هذا الاختبار يلعب التباين المقدر لخطأ التجريب دوراً رئيسياً كمقياس يقاس بالنسبة إليه الانحراف عن الخطية ، وهذا هو السبب في ضرورة أن تصمم التجارب التي تهدف إلى قياس جودة العلاقات الخطية بحيث يكون لكل قيمة س قيمتان أو أكثر من قيم ص وإلا ما استطعنا الحصول على هذا المقياس .

ملاحظة :

إذا وجدنا من الخطوة (أ) أن الانحدار الخطي ذو دلالة ووجدنا من الخطوة (ب) أن الانحراف عن الخطية غير ذي دلالة ، نحكم بأن العلاقة الخطية هي علاقة جيدة وتصف بمقدارة العلاقة الحقيقية بين المتغيرين . أما إذا وجدنا أن كلا من الانحدار الخطي والانحراف عن الخطية ذو دلالة فنحكم بأنه بالرغم من وجود علاقة خطية بين المتغيرين إلا أنها ليست أحسن العلاقات التي تعبر عن حقيقة العلاقة بينهما وعليها إذا أردنا أن نبحث عن علاقة أفضل .

مثال (٩ - ٩) :

ابحث خطية العلاقة بين الطول والوزن مستخدماً بيانات المثال (٩ - ٨) .

الحل :

( أولاً ) نبدأ بتحليل التباين للمتغير ص- للحكم على دلالة الاختلاف بين متوسطات المجموعات الصادية .

$$F = 5266,67 \quad \text{سبق إيجادها بالمثال (٩-٨)} \quad (28) \quad \text{الكلي}$$

$$٢٢ \text{ (بين المجموعات) } = \frac{١٧٠٠}{١٢} - \frac{٢٥١٥}{٣} + \frac{٢٩٥}{٢} + \frac{٥٢٥}{٤} + \frac{٣٦٥}{٣}$$

$$٣ = p \quad ٤٤٠٢,٠٨ =$$

$$٨ = p \quad ٨٦٤,٥٩ = ٤٤٠٢,٠٨ - ٥٢٦٦,٦٧ = \text{٢٢ (خطأ التجريب)}$$

$$\text{من (٢٥) : ف} \quad ١٣,٥٧٧ = \frac{٣ \div ٤٤٠٢,٠٨}{٨ \div ٨٦٤,٥٩}$$

وبما أن  $F_{[٨,٣], ٠,٠١} = ٧,٥٩$  نرفض الفرض الصفري عن تساوى المتوسطات عند مستوى الدلالة  $٠,٠١$  ونحكم بأن أوزان الرجال ليست مستقلة عن أطوالهم . ومادام الأمر كذلك نستمر في التحليل .

(ثانياً) نقوم بتحليل الانحدار للاختلاف بين المجموعات كما يلي :

$$٢٢ \text{ (الانحدار الخطى) } = ٤٣٤١,٨٧٠٨ \text{ سبق إيجاداه بالمثال (٨-٩) } ١ = p$$

$$\therefore ٢٢ \text{ (الانحراف عن الخطية) } = ٢٢ \text{ (بين المجموعات) } - ٢٢ \text{ (الانحدار الخطى)}$$

$$٤٣٤١,٨٧٠٨ - ٤٤٠٢,٠٨ =$$

$$٦٠,٢٠٩٢ = \quad ٢ = p$$

بضم هذه النتائج إلى نتائج الخطوة السابقة يتج جدول التباين الآتى :

الجدول (٩ - ٩)

مصدر التباين	د. ح	د. ح
بين المجموعات	٣	٤٤٠٧,٠٨
الانحدار الخطي	١	٤٣٤١,٨٧٠٨
الانحراف عن الخطية	٢	٦٠,٢٠٩٢
خطأ التجريب	٨	٨٦٤,٥٩
الكل	١١	٥٢٦٦,٦٧

(أ) اختبار وجود علاقة خطية

$$F_{٢٦} = \frac{١ \div ٤٣٤١,٨٧٠٨}{١٠ \div (٨٦٤,٥٩ + ٦٠,٢٠٩٢)} = ٤٦,٩٥$$

بما أن  $F_{١٠, ١١} = ١٠,٠٠$  نرفض الفرض الصفري أن  $\beta = ٠$  ونحكم بوجود علاقة خطية بين أطوال الرجال وأوزانهم .

(ب) اختبار دلالة الانحراف عن خط الانحدار

$$F_{٢٧} = \frac{٣٠,١٠٤٦}{١٠٨,٠٧٣٨} = ٠,٢٨$$

وإذن نقبل الفرض الصفري بعدم وجود انحراف ذي دلالة عن خط الانحدار .

## الخلاصة :

من هذه التجربة نستخلص أن العلاقة الخطية التى تمثلها معادلة الانحدار  
ش = - ١٧٩,٣٥ + ٥,٠٢٩ س تعبر بجدارة عن العلاقة الحقيقية بين أطوال وأوزان  
الرجال . ومع ملاحظة أن معامل التحديد

$$r^2 = \frac{[ \sum (س - \bar{س}) (ش - \bar{ش}) ]^2}{\sum (س - \bar{س})^2 \sum (ش - \bar{ش})^2} = \frac{(٨٦٣,٣٤)^2}{٥٢٦٦,٦٧ \times ١٧١,٦٦٧} = ٠,٨٢٤٤$$

نرى أن العلاقة الخطية تفسر حوالى ٨٢,٤٪ من الاختلاف الكلى فى قيم المتغير  
ش ، وهذا يدعم القول بخطية العلاقة بين المتغيرين .  
(٩ - ١٠) ملاحظات عن افتراضات الانحدار :

إن صحة الاستنتاجات الإحصائية تتوقف على مدى انطباق الافتراضات التى  
وضعت فى تعريف النموذج الذى بنيت عليه الدراسة على الموقف التجريبى الذى  
تتناوله هذه الدراسة . ولذلك ينبغى أن نفهم جيدا ما تعنيه هذه الافتراضات وما  
تتطلبه من شروط إجرائية تحسبا من العواقب التى قد تنجم عن وجود تناقضات  
ذات بال بين الافتراضات الموضوعية والظروف الفعلية لعملية التجريب . كما ينبغى  
أن نكون على استعداد لتغيير افتراض أو تعديله إذا اتضح لنا عدم إمكانية تحقيقه عمليا  
ولو بشئ من التقريب ، على أن نراعى ما يستلزمه هذا التغيير أو التعديل بالنسبة  
لما تقدمه من استنتاجات .

لنعتبر الافتراضات الثلاثة التى استخدمناها فى الانحدار الخطى البسيط ولنبدأ  
بالافتراض الأول . إن هذا الافتراض يتضمن أن يكون المتغير س متغيرا غير عشوائى

وهذا يعنى أن الباحث يتحكم تجريبيًا في هذا المتغير ويستطيع تسجيل القيم التي يدخلها في بياناته بدقة تامة . غير أنه في كثير من المواقف التجريبية لا يتحقق هذا الافتراض فقد يكون الباحث مهتمًا بالحصول على بيانات عن المتغير - دون أن يكون متحكمًا فيه أى بصرف النظر عن كون هذا المتغير عشوائيًا أو غير عشوائيًا مادامت هذه البيانات تعبر في نظر الباحث عن قيم نموذجية لهذا المتغير . وفي بعض الدراسات يضطر الباحث إلى اختيار مشاهدات تقع في مدى معين محدد من قبل أو يأخذ منها قيمة معينة محددة مسبقًا . في مثل هذه الحال ينبغي للباحث أن يستبدل بهذا الافتراض افتراضًا آخر يتلاءم مع الموقف الذي يتناوله ، كأن يضع الافتراض «المتغير - قيمه مشروطة» وهذا الافتراض الجديد لا يؤثر في سلامة استخدام طريقة المربعات الصغرى إلا أن الاستنتاجات الإحصائية التي نخرج بها عن المجتمع الذي ندرسه ينبغي أن تكون مشروطة بالنسبة للمتغير - أى لا يجوز تعميمها لقيم غير ممثلة في البيانات التي استخدمت في الدراسة أو ليس لها نفس خصائصها .

أما الافتراض الثاني فيعنى أن الصيغة الحقيقية للعلاقة بين المتغيرين هي الصيغة الخطية ، وعلى هذا لا تكون استنتاجاتنا صحيحة إلا إذا كان لدينا ما يضمن سلامة هذه الصيغة ، ويساعدنا في ذلك الأسلوب المذكور في نتيجة البند (٩ - ٥) لاختبار خطية هذه العلاقة كما يساعدنا الأسلوب المبين بالبند (٩ - ٩) لاختبار دلالة الانحراف عن خط الانحدار . فإذا تبين لنا عدم انطباق هذه الصيغة في موقف ما ينبغي أن نبحث عن صيغة أخرى تناسبه .

وبالنسبة للافتراض الثالث فهو يعنى أنه عند أى قيمة ثابتة من المتغير - تتصف القيم التي يأخذها المتغير - بما يلي :

(١) قيم - مستقلة ، أى أن صغر أو كبر الخطأ العشوائى في أحدها لا يؤثر في مقدار الخطأ في القيم الأخرى . وهذه الصفة يمكن تحقيقها عمليًا بإحكام عملية التجريب وخاصة فيما يتعلق بعشوائية العينة .

(٢) قيم  $\sigma$  لها تباين ثابت  $\sigma^2$  ، وهذه الصفة تتحقق تلقائيا إذا كانت هذه القيم هي مشاهدات مستقلة من نفس المجتمع .

(٣) قيم  $\sigma$  تتوزع توزيعا معتمدا .

وجدير بالإشارة هنا إلى أنه بصفة عامة تكون النماذج المستخدمة بما يصاحبها من فروض هي نماذج نظرية قلما تتحقق في الواقع العمل إلا على وجه التقريب . ونحن إذ نستخدم هذه النماذج تتعلق بالأمل في ألا تؤدي بنا هذه الحقيقة إلى وجود فروق كبيرة بين التقديرات التي نحصل عليها منها وبين القيم الحقيقية للبارامترات التي نبحث عنها ، كما تتعلق بالأمل بأن تكون الإجراءات التي اتخذناها في عملية التقدير هي إجراءات مناسبة حتى إذا كان الموقف الذي نتأوله لا يحقق الفروض تحقيقا تاما .

#### (٩ - ١١) استخدامات الانحدار :

لعله من المناسب الآن أن نبرز الأغراض التي يستخدم الانحدار الخطي البسيط من أجلها . ومن الدراسة التي مرت بنا في هذا الفصل يمكننا تلخيص هذه الأغراض فيما يلي :

(١) التنبؤ بقيم متغير عشوائي  $y$  بمعلومية متغير رياضي  $x$  مع تقدير درجة دقة هذا التنبؤ . هذا مع ملاحظة أن خط الانحدار هو خط مستقيم يمتد بغير نهاية من الطرفين . ولكننا في عملية التنبؤ لا يجوز التنبؤ بقيم صادية مناظرة لقيم سلبية نخرج كثيرا عن مدى القيم التي استخدمت في إنشاء هذا الخط إلا إذا كان لدينا ما يبرر ذلك . فمثلا إذا أوجدنا خط انحدار لأطوال الذكور بين العمر ١٠ والعمر ١٥ فمن الخطأ استخدام هذا الخط للتنبؤ بأطوال الرجال في أعمار فوق العشرين لأن الطول يتوقف عند بلوغ سن التضج .

(٢) دراسة طبيعة العلاقة بين متغيرين  $x$  ،  $y$  وشكل المنحنى أو الدالة التي تعبر عن هذه العلاقة .

(٣) تفسير بعض الاختلاف في قيم المتغير - بدلالة الاختلاف في قيم - على أساس اتخاذ المتغير - كضابط إحصائي statistical control .

(٤) دراسة ما إذا كان التغير في - يسبب ولو جزئيا التغير في - ، مع ملاحظة أن دراسة السببية تحتاج إلى أكثر من إثبات وجود علاقة ذات دلالة بين المتغيرين ، إذ أن هذه العلاقة قد تكون ناشئة عن وجود متغيرات أو عوامل أخرى تؤثر في المتغيرين - ، - معا . فمثلا في الأحياء المزدحمة من بعض المدن الكبيرة نجد علاقة ذات دلالة بين ازدحام السكان والتدنن الرئوى ( السل ) فهل هذا يعنى أن الازدحام سبب في الإصابة بالتدنن الرئوى ؟ لا نستطيع الإجابة بالإيجاب أو النفي عن هذا السؤال إذا لاحظنا أنه في الأحياء المزدحمة بصفة عامة يكون مستوى الحياة منخفضا وبالتالى يكون هناك سوء تغذية للسكان ، وقد يكون سوء التغذية هو السبب الحقيقى لهذا المرض . إن تقرير السببية أمر متروك للباحث يفتى فيه بما لديه من خبرة ومعلومات عن المتغيرات التى يتناولها مستعينا بما يستخلصه من التحليل الإحصائي .

(٥) يستخدم الانحدار أيضا في أنواع أخرى من التحليل الإحصائي منها تحليل التغيرات حيث يكون الهدف معرفة مدى تأثير عامل نوعى على متغير عددى - بعد استبعاد أثر متغير عددى - مرتبط بالمتغير - ، وهذا ما سنتناوله في فصل لاحق .

## تمارين (٩ - ٢)

( اعتبر أن افراضات الانحدار متوفرة ) .

(١) في تجربة عن تأثير طاقة التمثيل الغذائى بدرجة الحرارة على نوع من الطيور في خرة ضوئية ثابتة مدتها ١٠ ساعات ، اختيرت أربع درجات حرارة هي ٥ ، ١٠ ، ١٥ ، ٢٠ وعرضت خمسة طيور لكل من هذه الدرجات ثم حسبت مقادير طاقة التمثيل الغذائى لكل من هذه الطيور وسجلت بالجدول الآتى . أوجد معادلة



الانحدار الخطي لطاقة التمثيل الغذائي على درجة الحرارة واختبر جودة العلاقة الخطية بين هذين المتغيرين .

درجات الحرارة (س)				
٢٠	١٥	١٠	٥	
١٥,٨	١٨,٣	٢٤,٢	٢٣,٢	
١٥,٢	١٨,٨	٢٤,٢	٢٣,٥	
١٥,٦	١٨,٦	٢٤,١	٢٣,٤	طاقة التمثيل الغذائي (ص)
١٥,٣	١٨,٧	٢٤,٣	٢٣,١	
١٥,١	١٨,١	٢٤,٢	٢٣,٦	

لكل من العينات الثلاث الآتية أوجد معادلة الانحدار الخطي واختبر جودة العلاقة الخطية :

(٢)

س				
٦	٤	٢	١	٠
١١	٨	١	٠	١-
١٣	١٢	٥	٢	١

ص

(٣)

س			
٤	٣	٢	١
١٠	٩	٥	٠
١٤	١٣	٧	٢

ص

(٤)

٢٥	٢٠	١٥	١٠	٥	٢
٣٨٣	٣٤٥	٢٩٨	٢٤٢	١٧١	٦٠
٣٨٨	٣٥١	٣٠٤	٢٤٥	١٧٦	٦٤

ص

## الفصل العاشر

### الارتباط الخطى البسيط

### SIMPLE LINEAR CORRELATION

#### (١٠ - ١) الانحدار الخطى والارتباط الخطى :

في دراسة الانحدار الخطى لمتغير حقيقي  $y$  على متغير آخر  $x$  فرضنا أن العلاقة بينهما على الصورة  $y = \alpha + \beta x$  واستخرجنا من العينة أحسن تقديرين  $\hat{\alpha}$  ،  $\hat{\beta}$  للبارامترين المجهولين  $\alpha$  ،  $\beta$  في ضوء مبدأ المربعات الصغرى ، ومن ثم أوجدنا معادلة  $\hat{y} = 1 + 2x$  ، يمكننا من التنبؤ بأحسن قيمة للمتغير  $y$  عند قيمة معطاة للمتغير  $x$  ، واستخدمنا تحليل التباين لتفسير الانحدار ، كما خرجنا ببعض الاستنتاجات الإحصائية في صورة اختبارات دلالة وفترات ثقة . وهذا كله يدخل تحت الموضوع المسمى بتحليل الانحدار . ويقترن بهذا الموضوع موضوع آخر يسمى تحليل الارتباط وهو يهتم بالبحث عن عدد نقيس به درجة الاعتماد المتبادل بين المتغيرين ودقة العلاقة المفروضة بينهما ، فإذا فرضنا أن العلاقة بين المتغيرين خطية يكون اهتمامنا منصّباً على إيجاد عدد أو مقياس يعبر عن درجة جودة العلاقة الخطية في وصف العلاقة الحقيقية بين المتغيرين واختبار دلالة هذا المقياس .

#### (١٠ - ٢) افتراضات الارتباط الخطى البسيط :

تقتضي دراسة الارتباط بين متغيرين حقيقيين  $x$  ،  $y$  وضع افتراضات يختلف بعضها عن تلك التي وضعت لدراسة الانحدار . وسنضع هنا الافتراضات الآتية :

### الافتراض الأول :

« كل من المتغيرين  $x$  ،  $y$  هو متغير عشوائي » .  
فمثلا قد يعبر المتغيران عن طول ذراع الإنسان وطول رجله ، أو عن عمر الزوج وعمر الزوجة عند الزواج ، أو عن وزن الدجاجة وعدد البيض الذى تنتجه أو عن درجة الرياضيات ودرجة الفيزياء لمجموعة من الطلاب ..

### الافتراض الثانى :

« إن العلاقة بين المتغيرين  $x$  ،  $y$  هى علاقة خطية » بمعنى أن متوسط قيم  $y$  المناظرة لقيمة معينة  $x$  يأخذ الصورة

$$y = \alpha + \beta x \quad (1)$$

حيث  $\alpha$  ،  $\beta$  بارامتران مجهولان .

### الافتراض الثالث :

(أ) للمتغير  $x$  توزيع معتدل  
(ب) عند أى قيمة ثابتة  $x$  يكون للمتغير  $y$  توزيع معتدل متوسطه  $\alpha + \beta x$   
وتباينه عدد ثابت مجهول  $\sigma^2$  لا يتوقف على  $x$  .

وقد يكون من المفيد أن نشير إلى أن هذه الافتراضات تكافئ رياضيا القول بأن التوزيع المشترك للمتغيرين  $x$  ،  $y$  هو ذلك المسمى بالتوزيع المعتدل ذى المتغيرين

bivariate normal distribution

(١٠ - ٣) معامل الارتباط العزمى ( بيرسون ) :

**PRODUCT MOMENT CORRELATION COEFFICIENT**  
**(PEARSON 1900)**

تعريف :

إذا كان  $x$  ،  $y$  متغيرين عشوائيين ورمزنا بالرمز  $\sigma_x$  للانحراف المعيارى

للمتغير  $\sigma$  وبالرمز  $\sigma$  للانحراف المعياري للمتغير  $\sigma$  وبالرمز  $\sigma$  لتغاير  $\sigma$  (  $\sigma$  ،  $\sigma$  ) فإن العدد  $\sigma$  المعروف بالصيغة :

$$\frac{\sigma}{\sigma} = \sigma$$

يسمى بمعامل الارتباط العزمي بين  $\sigma$  ،  $\sigma$  .

يمكن رياضيا إثبات ما يلي :

(١) لا تزيد القيمة المطلقة للعدد  $\sigma$  عن الواحد الصحيح ، أى أن

$$-1 \leq \sigma \leq 1$$

(٢) إذا كان المتغيران  $\sigma$  ،  $\sigma$  مستقلين فإن  $\sigma = \sigma$  صفر غير أن العكس ليس من الضروري أن يكون صحيحا ، أى أنه إذا كان  $\sigma = \sigma$  . فليس من الضروري أن يكون المتغيران مستقلين بل قد يكون بينهما علاقة غير خطية . أما إذا كان التوزيع المشترك للمتغيرين  $\sigma$  ،  $\sigma$  هو التوزيع المعتدل ذو المتغيرين فإن انعدام معامل الارتباط يستلزم استقلال المتغيرين .

(٣) تكون هناك علاقة دالية خطية بين المتغيرين  $\sigma$  ،  $\sigma$  :

$$\sigma = a + b \sigma \quad \text{و} \quad \sigma = \bar{a} + \bar{b} \sigma$$

إذا وإذا فقط كان معامل الارتباط  $\sigma$  يساوى ١ أو - ١ .

وهذه الخاصة تعنى أنه إذا كان هناك علاقة خطية بين  $\sigma$  ،  $\sigma$  فإن ذلك ينعكس على قيمة  $\sigma$  فيجعلها مساوية للعدد ١ ( ونقول حينئذ أن هناك ارتباط تام موجب بين المتغيرين ) أو مساوية للعدد - ١ ( ونقول حينئذ أن هناك ارتباط تام سالب ) ، والعكس صحيح .

من هذا نرى أن معامل الارتباط العزمي ليس مقياسا للاعتماد المتبادل بين

المتغيرين بشكل عام وإنما هو مقياس لدرجة الاعتماد الخطي بينهما ، وينبغي أن نتذكر ذلك دائما .

ولتقدير قيمة  $r$  من عينة عشوائية :

$\{(s_1, v_1), (s_2, v_2), \dots, (s_n, v_n)\}$  نستخدم المقياس الآتي الذي يتركب بكل بساطة من التقديرات غير المتحيزة للقيم التي يتركب منها  $r$  :

$$(2) \quad \frac{\bar{E}}{\bar{E} \cdot \bar{E}} = r$$

حيث  $\bar{E} = \frac{1}{n} \sum (s_i - \bar{s})(v_i - \bar{v})$  تغاير  $(s, v)$

$$(3) \quad \frac{1}{1 - r} = \frac{\sum (s_i - \bar{s})^2 + \sum (v_i - \bar{v})^2}{n}$$

$\bar{E}' = \frac{1}{n} \sum (s_i - \bar{s})(v_i - \bar{v})$  تباين  $(s)$

$$(4) \quad \frac{1}{1 - r} = \frac{\sum (s_i - \bar{s})^2 + \sum (v_i - \bar{v})^2}{n}$$

$\bar{E}' =$  تباين  $(v)$

ملاحظة ( ١ )

من السهل إثبات أن (٢) يمكن أن تكتب على الصورة

$$(5) \quad r = \frac{\sum (s_i - \bar{s})(v_i - \bar{v})}{\sqrt{[\sum (s_i - \bar{s})^2][\sum (v_i - \bar{v})^2]}}$$

وهذه الصيغة أفضل من الصيغة (٢) من الناحية الحسابية .

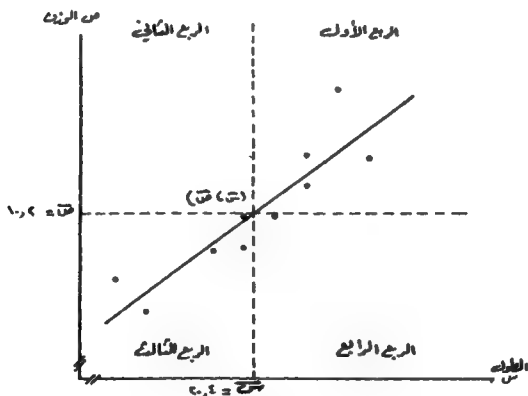
وقبل أن نوضح خصائص معامل الارتباط  $r$  والحكمة في أخذه كمقياس للارتباط الخطي نتناول المثال الآتي :

مثال (١٠ - ١) :

أخذت عينة عشوائية من نوع معين من نبات البسلة وقيست أطوالها بالمليمتر (س) وأوزانها بالمليجرام (ص) فوجد ما يلي :

س :	١٧	٢٢	٢٠	٢٣	٢٤	٢٢	٢٠	٢١	١٦
ص :	٧	١٢	٩	١٤	١٢	١١	١٠	٩	٨

ارسم شكل الانتشار لاستيضاح خطية العلاقة بين المتغيرين ثم أوجد معامل الارتباط العزمي .



الشكل (١٠ - ١) شكل الانتشار لأطوال وأوزان عينة من ١٠ نباتات بسلة

يوحى شكل الانتشار بأن التخطئ تميل إلى أن تقع على خط مستقيم موجب الميل ، وهذا يشير مبدئيا إلى خطية العلاقة بين المتغيرين .

الجدول (١٠ - ١)

إيجاد معامل الارتباط العزمي من بيانات المثال (١٠ - ١)

ص	ص	ص ص	ص	ص
١٧	٧	١١٩	٢٨٩	٤٩
٢٢	١٢	٢٦٤	٤٨٤	١٤٤
٢٠	٩	١٨٠	٤٠٠	٨١
٢٣	١٤	٣٢٢	٥٢٩	١٩٦
٢٤	١٢	٢٨٨	٥٧٦	١٤٤
٢٢	١١	٢٤٢	٤٨٤	١٢١
٢٠	١٠	٢٠٠	٤٠٠	١٠٠
١٩	٩	١٧١	٣٦١	٨١
٢١	١٠	٢١٠	٤٤١	١٠٠
١٦	٨	١٢٨	٢٥٦	٦٤
٢٠٤	١٠٢	٢١٢٤	٤٢٢٠	١٠٨٠

معامل الارتباط العزمي هو ( بالتعويض في ٥ )

$$= \frac{(2124 \times 10) - (102 \times 204)}{}$$

$$= \frac{[ (102) - 1080 \times 10 ] [ (204) - 4220 \times 10 ] \sqrt{}}$$

$$= \frac{432}{}$$

$$= \frac{432}{}$$

$$= \frac{432}{410,899}$$

$$= \frac{432}{396 \times 0.84 \sqrt{}}$$

$$= 0.898 \text{ تقريباً .}$$



#### (١٠ - ٤) مميزات معامل الارتباط العزمي .

##### (١) تركيب معامل الارتباط :

إن قدرة معامل الارتباط المعرف في (٢) على تقدير درجة العلاقة الخطية بين المتغيرين  $s$  ،  $\bar{s}$  تبين من الدراسة الرياضية للنموذج الذى وضعت له الافتراضات المذكورة آنفاً . غير أننا نستطيع أن نرى ذلك بطريقة بسيطة كالآتي .

إذا تأملنا نقط مستقيم موجب الميل ، لاحظنا أنه كلما كان الإحداثي السيني لنقطة كبيراً كلما كان إحداثيها الصادي كبيراً أيضاً . أما إذا كان المستقيم سالب الميل فإنه كلما زاد الإحداثي السيني كلما نقص الإحداثي الصادي . وعلى ذلك فإن قياس خطية العلاقة بين المتغيرين تتطلب مقياساً حساساً لدرجة اقتران القيم الكبيرة للمتغير  $s$  بالقيم الكبيرة للمتغير  $\bar{s}$  إذا كانت العلاقة موجبة ، ولدرجة اقتران القيم الكبيرة للمتغير  $s$  بالقيم الصغيرة للمتغير  $\bar{s}$  إذا كانت العلاقة سالبة وهذه الحساسية نجدها في التغاير  $\bar{c}_s$  ففي المثال (١٠ - ١) السابق ، إذا رسمنا الخط الرأسى  $s = \bar{s} = ٢٠,٤$  والخط الأفقى  $s = \bar{s} = ١٠,٢$  في شكل الانتشار لاحظنا أنه في أغلب الحالات بل في جميع الحالات ما عدا حالة واحدة هى حالة النقطة (٢١ ، ١٠) الواقعة في الربع الرابع ما يلى :

حين تكون  $s$  كبيرة ( أكبر من الوسط الحسابى  $\bar{s}$  ) فإن  $\bar{s}$  تكون المناظرة تكون كبيرة أيضاً ( أكبر من الوسط الحسابى  $\bar{s}$  ) وحين تكون  $s$  صغيرة ( أصغر من  $\bar{s}$  ) فإن  $\bar{s}$  المناظرة تكون صغيرة أيضاً ( أصغر من  $\bar{s}$  ) . أى أن :

حين  $(s - \bar{s})$  موجبة تكون  $(\bar{s} - \bar{s})$  موجبة وحين  $(s - \bar{s})$  سالبة تكون  $(\bar{s} - \bar{s})$  سالبة وفي كلتا الحالتين تكون  $(s - \bar{s})(\bar{s} - \bar{s})$  موجبة ويكون المجموع  $\sum (s - \bar{s})(\bar{s} - \bar{s})$  موجبا .

وبالتالى يكون التغير  $\bar{E}_s$  موجبا (إلا إذا كانت قيمة  $(s - \bar{s})$  ص - ص) للنقطة (٢١ ، ١٠) كبيرة جداً وهذا لم يحدث .

ويلاحظ أن قيمة  $\bar{E}_s$  في هذا المثال كبيرة لأنها متوسط مجموع ٩ أعداد موجبة وعدد واحد سالب . ولو كانت النقطة التي في الربع الرابع قد وقعت في الربع الأول أو الثالث لزادت قيمة  $\bar{E}_s$  وبالعكس لو كانت إحدى النقط الواقعة في الربعين الأول أو الثالث قد وقعت في الربع الثاني أو الرابع لنقصت قيمة  $\bar{E}_s$

وعلى ذلك فإن التغير يعكس أمرين هما : درجة جودة العلاقة الخطية واتجاه هذه العلاقة . ولهذا يركز معامل الارتباط  $r$  المعروف في (٢) على التغير  $\bar{E}_s$  . هذا مع ملاحظة أن هذا المعامل يكون موجبا أو سالبا بحسب كون العلاقة موجبة أو سالبة .

غير أن قيمة التغير تعتمد على حجم العينة  $n$  كما هو واضح من التعريف (٣) ، كما أنها تعتمد على وحدات القياس . ولكي يكون مقياس الارتباط عاماً ينبغي أن يكون مستقلاً عن حجم العينة وعن وحدات القياس . ولذلك ينبغي تعديل التغير ليحقق هذين الشرطين قبل أخذه كمقياس للارتباط ، وهذا ما يحققه قسمة التغير  $\bar{E}_s$  على حاصل ضرب الانحراف المعياري  $\bar{E}_s$  للمتغير  $s$  والانحراف المعياري  $\bar{E}_s$  للمتغير  $s$  . وهذا الإجراء يكافئ إيجاد تغير  $s$  ،  $s$  بعد وضع القيم في الصورة

$$\frac{s - \bar{s}}{\bar{E}_s} , \frac{\bar{s} - \bar{s}}{\bar{E}_s}$$

بهذا الإجراء يكون المقياس .

$r = \frac{\bar{E}_s}{\bar{E}_s \cdot \bar{E}_s}$  هو مقياس مطلق ( لا يعتمد على وحدات القياس ) ولا يعتمد على حجم العينة .

(ب) من مزايا معامل الارتباط أنه متماثل في  $r$  ،  $r$  بمعنى أن قيمته لا تتغير بوضع  $r$  ،  $r$  كل مكان الآخر .

(ج) قيمة معامل الارتباط  $r$  مستقلة عن اختيار نقطة الأصل لأن كلا من التباين والانحراف المعياري يتمتع بهذه الصفة ، وهذا يعني أن قيمة  $r$  لا تتغير بإضافة أو طرح عدد ثابت من جميع قيم  $r$  أو بإضافة أو طرح عدد ثابت من جميع قيم  $r$  . ونستفيد من هذه الخاصية أحياناً في تبسيط حساب قيمة  $r$  .  
(د) يمكن إثبات أن القيمة المطلقة للمقياس  $r$  لا تزيد عن الواحد الصحيح .  
أي أن .

$$(٦) \quad -1 \leq r \leq 1$$

ولما كانت  $r$  مقياساً لدرجة الارتباط الخطي بين متغيرين فإننا نقول إن المتغيرين غير مرتبطين خطياً إذا كانت  $r = 0$  ( وهذا لا يمنع من وجود ارتباط من نوع آخر ) . وكلما اقتربت  $r$  من الواحد كلما أوحى ذلك بوجود علاقة خطية ( موجبة أو سالبة ) بين المتغيرين .

ملاحظة ( ٢ )

هناك علاقة مفيدة بين الخطأ المعياري  $\sigma_r$  المعروف بالمعادلة (٨) بالبند (٩ - ٤) في موضوع الانحدار الخطي البسيط ومعامل الارتباط  $r$  وهي :

$$(٧) \quad \sigma_r = \frac{1 - r^2}{n - 2} \sigma_{\text{ع}} \quad (1 - r^2)$$

ففي المثال (١٠ - ١) نستطيع إيجاد الخطأ المعياري كالتالي :

$$\sigma_r = \frac{1}{n - 2} [\text{ع}(\text{ص}) - \text{ع}(\text{ص})^2] = \frac{1}{10} (1.02 - 10.80) = 0.98$$

$$\text{إذن } \bar{C}_{\text{م.س}} = \frac{9}{8} \times 4,4 (1 - 0,898) = 0,9083$$

$$\text{إذن } \bar{C}_{\text{م.س}} = 0,9789 \text{ تقريباً .}$$

(١٠ - ٥) دلالة معامل الارتباط العزمي :

ليكن  $\bar{C}_{\text{م.س}}$  هو معامل الارتباط الخطي بين المتغيرين  $\bar{C}_{\text{م.س}}$  ،  $\bar{C}_{\text{م.س}}$  في المجتمع ،  $\bar{C}_{\text{م.س}}$  هو معامل الارتباط الناتج من العينة . تحت الافتراضات الثلاثة المذكورة نستطيع اختبار أى فرض عن القيمة الحقيقية للمعامل  $\bar{C}_{\text{م.س}}$  ووضع حدود ثقة لهذا المعامل .

(أولاً) اختبار الفرض  $\bar{C}_{\text{م.س}} = \text{صفر}$  .

لاختبار الفرض الصفري  $\bar{C}_{\text{م.س}} = \text{صفر}$  ( المتغيران غير مرتبطين خطياً )

ضد الفرض الآخر  $\bar{C}_{\text{م.س}} \neq \text{صفر}$

نستخدم الإحصاءة

$$(٨) \quad T = \frac{\sqrt{n-2} \bar{C}_{\text{م.س}}}{\sqrt{1 - \bar{C}_{\text{م.س}}^2}}$$

التي لها توزيع ت بدرجات حرية  $n - 2$  بشرط صحة الفرض الصفري .

مثال (١٠ - ٢) :

في المثال (١٠ - ١) اختبر ما إذا كان المتغيران  $\bar{C}_{\text{م.س}}$  ،  $\bar{C}_{\text{م.س}}$  غير مرتبطين خطياً

مستخدماً مستوى الدلالة ٠,٠١

الحل :

$$\text{لدينا } n = 10 , r = 0,898$$

ف : م = ٠ ، ف : م ≠ ٠ (اختبار ذو جانين)

$$\text{من (٨) : } T = \frac{\sqrt{2-10} \cdot ٠,٨٩٨}{\sqrt{٠,٨٩٨ - ١٧}} = ٥,٧٧٣$$

من الجدول  $T_{(٨), ٠,٠١} = ٣,٣٥٥$

بما أن  $٣,٣٥٥ < ٥,٧٧٣$  نرفض الفرض الصفري عند مستوى الدلالة ٠,٠١ ونحكم بوجود ارتباط خطي بين المتغيرين .

### ملاحظة (٣)

الإحصاءة (٨) التي تختبر الفرض م = ٠ ضد الفرض م ≠ ٠ تكافئ الإحصاءة (١٠) بالبند (٩ - ٥) في موضوع الانحدار الخطي البسيط عند استخدامها لاختبار الفرض  $\beta = ٠$  ضد الفرض  $\beta \neq ٠$  . وذلك لأن كلاهما يقيس وجود أو عدم وجود علاقة خطية بين المتغيرين م ، م . وقد يبدو ذلك غريباً لأن الإحصاءة (٨) تتطلب أن يكون كلا المتغيرين معتمداً بينما الإحصاءة (١٠) تتطلب أن يكون المتغير م فقط معتمداً . وتزول هذه الغرابة إذا علمنا أن فيشر قد أثبت أنه في الحالة الخاصة التي يكون فيها م = ٠ فإن توزيع المعاينة لمعامل الارتباط م لا يتغير سواء كان المتغير م معتمداً أو غير معتمداً طالما كان المتغير م معتمداً .

ولذلك فإن الإحصاءة (٨) تصلح فقط لاختبار الفرض الصفري م = ٠ . ولا تصلح لاختبار أى فرض صفري آخر مثل م = ٥,٥ ، أو لاجداد فترات الثقة لمعامل الارتباط م في المجتمع ، وذلك لأنه حين يكون المتغيران مرتبطين ، أى حين م ≠ ٠ يكون توزيع م ملتوياً وبعيداً عن الاحتمالية . فكيف نختبر الفرض م = ٠ له حيث  $٠ \neq ٠$  ؟

(ثانياً) اختبار الفرض  $H_0 = E$  حيث  $E \neq 0$  .

وجد فيشر أنه عندما  $H_0 \neq 0$  فإن التحويل

$$(9) \quad \frac{1}{\sigma} = E \text{ لو } \frac{r+1}{r-1} , \quad \frac{1}{\sigma} = E_r \text{ لو } \frac{r+1}{r-1}$$

يعرف متغيراً عشوائياً  $E$  له توزيع معتدل على وجه التقريب متوسطه  $E$  وتباينه

$$(10) \quad \frac{1}{\sigma^2} = \frac{1}{E^2}$$

وأن هذا التوزيع يقترب بسرعة من التوزيع المعتدل بزيادة حجم العينة . وبذلك يكون للمتغير

$$(11) \quad \frac{E - E}{\sigma} = \text{ص}$$

توزيع معتدل قياسي على وجه التقريب، وبالتالي يجوز تناوله باستخدام جدول المساحات أسفل المنحنى المعتدل القياسي . كما أن الفترة

$$(12) \quad (E - E) \times \frac{\sigma}{\sqrt{r}} , \quad (E_r - E) \times \frac{\sigma}{\sqrt{r}}$$

تكون فترة ثقة بدرجة  $(1 - \alpha)$  للبارامتر  $E$  ، حيث  $\frac{\sigma}{\sqrt{r}}$  هي قيمة المتغير المعتدل القياسي  $\text{ص}$  التي تحقق المعادلة

$$L = \left( \frac{\sigma}{\sqrt{r}} < \text{ص} < \frac{\sigma}{\sqrt{r}} \right) = 1 - \alpha$$

ومن الفترة (١٢) نستطيع إيجاد فترة الثقة لمعامل الارتباط ص للمجتمع من التحويل (٩) .

ونتجنب مشقة حساب التحويل (٩) وُضع جدول يحول ص إلى ع ( بصرف النظر عن الإشارة ) هو الجدول (١١) بملحق هذا الكتاب كما وضع جدول يحول ع إلى ص هو الجدول (١٢) .

مثال (١٠ - ٣) :

في عينة عشوائية من ٢٨ زوجا من مجتمع معتدل ذى متغيرين وجد أن معامل الارتباط ٠,٧ هل هذه القيمة تتماشى مع الفرض القائل أن معامل الارتباط في المجتمع هو ٠,٥ ؟ أوجد فترة ثقة بدرجة ٩٥٪ لهذا المعامل .

الحل :

الفرض الصفري : ص = ٠,٥ الفرض الآخر ص ≠ ٠,٥

نحول كلا من معامل الارتباط في العينة ومعامل الارتباط المفروض للمجتمع إلى القيمتين المناظرتين لهما باستخدام الجدول (١١) .

$$ص = ٠,٧ \text{ تنتج } ع = ٠,٨٦٧$$

$$ص = ٠,٥ \text{ تنتج } ع = ٠,٥٤٩$$

$$\text{الخطأ المعياري} = \frac{1}{\sqrt{٢٥}} = \frac{1}{\sqrt{٣ - ٠,٧}} = ٠,٢$$

على أساس صحة الفرض الصفري نجد أن :

$$\therefore \text{ص} = \frac{٠,٥٤٩ - ٠,٨٦٧}{٠,٢} = ١,٥٩$$

وبما أن ١,٥٩ أقل من القيمة الحرجة ١,٩٦ لا يكون لدينا دليل يدعونا لرفض الفرض الصفري ، ونحكم بأن معامل الارتباط في المجتمع يمكن أن يكون ٠,٥ من الصيغة (١٢) ، نوجد فترة الثقة بدرجة ٩٥٪ للبارامتر ع كالآتي :

الحد الأدنى للفترة  $= 0,867 - 0,2 \times 1,96 = 0,475 = 0,48$  تقريباً

الحد الأعلى للفترة  $= 0,867 + 0,2 \times 1,96 = 1,209 = 1,26$  تقريباً

∴ فترة الثقة للبارامتر  $\bar{r}$  هي  $(0,48, 1,26)$ .

من الجدول (١٢) نجد أن  $(0,446, 0,851)$  هي فترة الثقة بدرجة ٩٥٪ لمعامل الارتباط.

مثال (١٠ - ٤) :

في عينة من ٩ أزواج من المشاهدات وجد أن معامل الارتباط  $-0,889$  أوجد فترة ثقة بدرجة ٩٩٪ لمعامل الارتباط في المجتمع.

الحل :

من الجدول (١١) :  $r = -0,889$  وإذن  $E = 1,417$  (مع إهمال إشارة  $r$ )

$$\text{الخطأ المعياري} = \frac{1}{\sqrt{3-2\sqrt{2}}} = \frac{1}{\sqrt{6}} = 0,408$$

الحد الأدنى للفترة للبارامتر  $\bar{r}$   $= 1,417 - 0,408 \times 2,58 = 0,366$ .

الحد الأعلى للفترة للبارامتر  $\bar{r}$   $= 1,417 + 0,408 \times 2,58 = 2,468$

∴ الفترة  $(0,37, 2,47)$  هي فترة ثقة بدرجة ٩٩٪ للبارامتر  $\bar{r}$ . من الجدول

(١٢) ومع استرجاع إشارة  $r$  تكون الفترة  $(-0,987, -0,354)$  هي فترة ثقة بدرجة ٩٩٪ لمعامل الارتباط  $r$ .



( ثالثاً ) اختبار دلالة الفرق بين معامل ارتباط عيتين مستقلتين :

نفرض أن لدينا عيتين مستقلتين حجمهما  $n_1$  ،  $n_2$  من الأزواج أعطيتا معامل ارتباط  $r_1$  ،  $r_2$  نريد أن نختبر ما إذا كان من الممكن اعتبار هاتين العيتين مأخوذتين من نفس المجتمع ( أو من مجتمعين لهما نفس معامل الارتباط ) ، أى نريد أن نختبر ما إذا كان الفرق بين  $r_1$  ،  $r_2$  ذا دلالة .

نحول القيمتين  $r_1$  ،  $r_2$  إلى القيمتين المناظرتين  $E_1$  ،  $E_2$  باستخدام الجدول ( ١١ ) . إذا كانت العيتان مستقلتين ومأخوذتين من نفس المجتمع المعتدل فإن الفرق بين  $E_1$  ،  $E_2$  يكون له توزيع معتدل تباينه يساوى مجموع تباينهما ، أى

$$\sigma_{E_1 - E_2}^2 = \frac{1}{n_1 - 3} + \frac{1}{n_2 - 3} \quad (١٣)$$

ويكون الخطأ المعياري للفرق  $|E_1 - E_2|$  هو الجذر التربيعي لهذا المقدار .  
وعلى ذلك نستطيع اختبار دلالة الفرق بين  $E_1$  ،  $E_2$  ( أى بين  $r_1$  ،  $r_2$  ) بواسطة التوزيع المعتدل .

مثال ( ١٠ - ٥ ) :

فى عينة عشوائية من ٥ أبقار وجد أن معامل الارتباط بين الزيادة فى الوزن ومقدار الغذاء المأكول ٠,٨٧ ، وفى عينة عشوائية مستقلة حجمها ١٢ بقرة وجد أن معامل الارتباط ٠,٥٦ ، فهل هذان المعاملان مختلفان اختلافاً جوهرياً ؟ استخدم مستوى الدلالة ٥% .

الحل :

الفرض الصفري  $H_0 =$  والفرض الآخر  $H_1 \neq$

من الجدول ( ١١ ) نجد أن  $r_1 = ٠,٨٧$  ومنها  $E_1 = ١,٣٣٣$

،  $r_2 = ٠,٥٦$  ومنها  $E_2 = ٠,٦٣٣$

تباين الفرق ع - ع يساوى مجموع التباينين

$$0.611 = \frac{1}{4} + \frac{1}{4} =$$

$$\therefore \text{الخطأ المعيارى للفرق} = \sqrt{0.611} = 0.782$$

على أساس صحة الفرض الصفرى ص = ص (أى ع = ع)

$$\text{ص} = \frac{(1.333 - 0.633) - \text{صفر}}{0.782} = 0.895$$

وهذه القيمة تقل عن ١,٩٦ فهى ليست ذات دلالة ولا يسعنا ألا أن نحكم بأن ص = ص .

#### (١٠ - ٦) التمييز بين الانحدار والارتباط فى دراسة المشكلات :

كل من الانحدار والارتباط الخطى البسيط يتناول العلاقة الخطية بين متغيرين كميين ص ، ص . إلا أن هناك مواقف تستدعى استخدام أسلوب الانحدار ولا يجوز أن تُدرس بأسلوب الارتباط ، كما أن هناك مواقف تستدعى استخدام أسلوب الارتباط ولا يجوز أن تُدرس بأسلوب الانحدار . ذلك أن الافتراضات التى يبنى عليها التحليل تختلف فى الانحدار عنها فى الارتباط خاصة فيما يتعلق بنوعية المتغير ص ، حيث نفترض فى الانحدار أن ص متغير رياضى لا يتأثر بالعوامل العشوائية بينما نفترض فى الارتباط أنه عشوائى . وهذا الاختلاف فى النظر إلى المتغير ص يعنى من الناحية العملية اختلافا فى طريقة المعالجة . وعلى ذلك فإن اختيار الأسلوب الذى يناسب مشكلة ما - انحدار أو ارتباط - يتوقف على الطريقة التى تتبع فى عملية المعالجة أى فى عملية جمع البيانات .

فلاستخدام أسلوب الانحدار يتطلب الأمر أن يختار الباحث قيما ثابتة من المتغير ص يحددها قبل إجراء التجربة ثم يقوم بملاحظة ما يظهر من القيم المناظرة للمتغير

من عند إجراء التجربة . فمثلا في دراسة العلاقة بين جرعات دواء مهدىء وقدرة الانسان على حل المشاكل المنطقية يبدأ الباحث بتحديد بضعة جرعات مختلفة التركيب من هذا الدواء ( قيم المتغير من ) ثم يختار عينة عشوائية من الأفراد يقسمها عشوائيا إلى مجموعات عددها هو عدد الجرعات المختلفة ويعطى لأفراد كل مجموعة واحدة من تلك الجرعات ، ثم يستخدم أحد الاختبارات لقياس القدرة المنطقية لكل فرد فيحصل على قيم للمتغير من . كذلك في دراسة العلاقة بين طول نبات ومحتوى النيتروجين في التربة يبدأ الباحث بتحديد عدة أحواض زراعية تختلف في محتوى النيتروجين ( قيم من ) ويوزع النبات في كل منها ثم يقيس أطوال النباتات بعد فترة من الزمن ليحصل على قيم للمتغير من . في مثل هاتين الحالتين تكون قيم من فقط خاضعة للمؤثرات العشوائية ، أما قيم من فتكون قيما ثابتة . وينبغي هنا تناول البيانات بأسلوب الانحدار حيث نقوم بإيجاد معادلة تشير إلى مدى اعتماد المتغير من على المتغير من ، ويحق لنا أن نصف من بأنه المتغير المستقل وأن نصف من بأنه المتغير التابع .

أما في أسلوب الارتباط فيتطلب الأمر أن يبدأ الباحث باختيار عينة عشوائية من المجتمع ثم يقوم بقياس كل من قيم من ، من لكل وحدة من وحدات العينة وبذلك تكون جميع القيم السينية والصادية خاضعة للمؤثرات العشوائية إذ يكون لكل منها حرية تامة في اتخاذ أى قيمة من القيم الممكنة في المجتمع . فمثلا في دراسة العلاقة بين طول الذراع وطول الرجل لمجتمع من الأطفال ، يختار الباحث عينة عشوائية من الأطفال ثم يقوم بقياس كل من المتغيرين . كذلك في دراسة العلاقة بين محتوى الكلسترول في الدم ووزن الجسم في مجتمع من المرضى بمرض معين نأخذ عينة عشوائية من هؤلاء المرضى ثم نقيس كلا من هذين المتغيرين . في مثل هاتين الحالتين تكون قيم كل من المتغيرين من ، من عشوائية . وينبغي هنا استخدام أسلوب الارتباط حيث نقوم بإيجاد عدد نقدر به الدرجة التي يتغير بها المتغيران معا دون تمييز بين متغير مستقل وآخر تابع .

وفي الحالات التي تتطلب دراستها استخدام أسلوب الانحدار يمكن من الناحية الحسابية إيجاد معامل الارتباط ، ولكن المعامل الناتج يكون مجرد عدد لا معنى له ولا يجوز أن يؤخذ كتقدير للارتباط بين المتغيرين . ( وهذا لا يتناقض مع استخدامنا لمربع معامل الارتباط وهو ما سميناه بمعامل التحديد ، في تحليل الانحدار كما رأينا في الفصل السابق ) .

كذلك ، في الحالات التي تتطلب دراستها استخدام أسلوب الارتباط يمكن من الناحية الحسابية إيجاد معادلة الانحدار ، ولكن المعادلة الناتجة لا تحقق الهدف منها كمعادلة تنبؤ أو لبيان العلاقة الخطية بين المتغيرين لأن التقديرين  $\alpha$  ،  $\beta$  ب يكونان في هذه الحالة تقديرين متحيزين للبارامترين  $\alpha$  ،  $\beta$  . ويتضح ذلك عندما نذكر أننا في قياس دقة المقدار  $\beta$  في تقدير البارامتر  $\beta$  نستخدم الخطأ المعياري  $\frac{\sqrt{MSE}}{n}$  الذي يعتمد على قيمة  $MSE = (SS - SS') / (n - 2)$  - راجع البند ( ٩ - ٥ - أولاً ) فإذا لم تكن قيم  $\alpha$  ثابتة فإن قيمة هذا الخطأ المعياري تختلف من عينة إلى أخرى وبالتالي تختلف درجة دقة التقدير من عينة إلى أخرى .

إن التمييز بين الحالات التي تدرس بأسلوب الانحدار وتلك التي تدرس بأسلوب الارتباط هو أمر على درجة كبيرة من الأهمية . وقد لوحظ أن الأمر يختلط في كثير من الحالات فمعامل مشكلات الانحدار على أنها ارتباط وتعامل مشكلات الارتباط على أنها انحدار ، بل ويصل الأمر أحياناً إلى معاملة المشكلة الواحدة على أنها انحدار وارتباط في آن واحد ، رغم أن طريقة المعاينة لا تسمح إلا باستخدام واحد فقط من هذين الأسلوبين . ويبدو أن السبب في هذا اللبس يرجع إلى وجود علاقات رياضية كثيرة بين المعاملات في هذين النوعين من التحليل ، غير أن العلاقات الرياضية شيء والتحليل الإحصائي شيء آخر .

وجدير بالذكر أنه من الممكن دراسة الانحدار حين يكون كل من المتغيرين  $\alpha$  ،  $\beta$  .

صه عشوائيا ، ولكن ذلك يتطلب استخدام نموذج إحصائي يختلف عن النموذج الذى استخدمناه فى الفصل السابق بحيث يسمح بوجود خطأ عشوائى فى المتغير صه . وسوف لا نتعرض للدراسة هذا النموذج خاصة وأنه يفضل دراسة مثل هذه الحالة بأسلوب الارتباط .

( ١٠ - ٧ ) معامل ارتباط الرتب ( سيرمان ) :

### RANK CORRELATION COEFFICIENT

بالرغم من أن معامل ارتباط بيرسون يستخدم أساسا فى حالة المتغيرات المتصلة إلا أنه يستخدم أيضا فى حالات أخرى . ومن هذه الحالات الحالة التى يتعذر فيها قياس المفردات بالمقياس العددي المعتاد وإنما يمكن قياسها بميزان الترتيب حيث تأخذ كل مفردة ترتيبين س ، ص تعبر الأولى عن ترتيب المفردة بالنسبة للمتغير الأول وتعبر الثانية عن ترتيبها بالنسبة للمتغير الثانى ويكون المطلوب قياس الارتباط بين الترتيبين . ومثال ذلك قيام اثنين من الحكام بترتيب عدد من المتقدمين لشغل وظيفة بحسب أفضليتهم لهذه الوظيفة ثم قياس مدى الاتفاق بين الحكمين . فى هذه الحالة يمكن إثبات أن معامل الارتباط الغزى يأخذ الصيغة البسيطة الآتية التى تعزى إلى سيرمان ويرمز له بالرمز  $r_{sr}$  حيث

$$r_{sr} = 1 - \frac{6 \sum F^2}{n(n^2 - 1)} \quad \text{وحيث } F_{sr} = S_{sr} - \bar{S} \bar{S} \quad (١٤)$$

مثال ( ١٠ - ٣ ) :

كانت تراتيب ١٢ طالباً فى مادتي الفيزياء س والرياضيات ص كما يلى :

س .	١١	٧	٩	١٢	١	٤	١٠	٣	٨	٥	٢	٦
ص .	١٠	٥	٨	١١	١	٦	١٢	٣	٧	٤	٢	٩

أوجد معامل الارتباط الخطى .

الحل :

بما أن المتغيرين مقياسان بميزان الترتيب فمن الأسهل استخدام معامل ارتباط الرتب كما في الجدول الآتي ، الذي يستخدم الصيغة (١٤) .

الجدول (١٠-٢)

إيجاد معامل ارتباط الرتب من بيانات الكمال (١٠-٣)

س	ص	ف	ف'
١١	١٠	١	١
٧	٥	٢	٤
٩	٨	١	١
١٢	١١	١	١
١	١	٠	٠
٤	٦	٢-	٤
١٠	١٢	٢-	٤
٣	٣	٠	٠
٨	٧	١	١
٥	٤	١	١
٢	٢	٠	٠
٦	٩	٣-	٩
		صفر	٢٦

بالتعويض في (١٤) نجد أن :

$$r = 1 - \frac{26 \times 6}{143 \times 12} = 0.91$$

ونحصل على هذه النتيجة بالضبط إذا استخدمنا معامل ارتباط بيرسون المعروف في (٥) ، مع ملاحظة أنه قد يوجد فرق طفيف يعود إلى عمليات تقريب الأعداد .

(١٠ - ٨) سميزات معامل ارتباط الرتب :

كما سبق القول ، يصلح معامل ارتباط الرتب المعروف في (١٤) لقياس الارتباط بين المتغيرات التي تقاس بميزان الترتيب وهو يصلح أيضاً للمتغيرات التي تقاس بالمقياس العددي المعتاد ولكن يكون من المرغوب فيه تعيين رتب لكل مفردة بدلا من قيمها العددية . ومن أهم سميزات هذا المعامل أنه لا يشترط فرض اعتدالية المتغيرين  $r$  ،  $s$  ، بل لا يشترط فرض وجود علاقة خطية بينهما ومن هنا فهو يفضل معامل الارتباط العزمي في الحالة التي لا يتوفر فيها هذان الشرطان ، وقياس الارتباط بواسطة معامل الرتب يعتبر من فصيلة المقاييس غير البارامترية أى التي لا تستلزم وضع فروض معينة على توزيعات المجتمعات .

ومن الأمثلة التي يستخدم فيها هذا المعامل في العلوم البيولوجية قياس الارتباط بين ترتيب انبثاق يرقات مجموعة من الحشرات وترتيب حجوماها ، أو بين ترتيب استنبات مجموعة من النباتات وترتيب تزهرها ، أو لقياس الاتفاق بين اثنين من البيولوجيين في ترتيبهما لمجموعة من الكائنات العضوية من حيث أكثرها شبهاً لصيغة معينة إلى أقلها شبهاً بها .

مثال (١٠ - ٤) :

في عشرة أنواع من السجائر وجدت المقادير الآتية بالمليجرام من القار والنيكوتين . حول هذه المقادير إلى رتب ثم أوجد معامل ارتباط الرتب لقياس درجة العلاقة بين محتويات القار والنيكوتين .

النوع : (١) (٢) (٣) (٤) (٥) (٦) (٧) (٨) (٩) (١٠)

محوى القسار : ١٤ ١٧ ٢٨ ١٧ ١٦ ١٣ ٢٤ ٢٥ ١٨ ٣١

محوى النيكوتين : ٠,٩ ١,١ ١,٦ ١,٣ ١,٠ ٠,٨ ١,٥ ١,٤ ١,٢ ٢,٠

الحل :

نرتب كلا من مفردات المجموعتين بحسب الأصغر فالأكبر (أو العكس) كما في الجدول الآتي ، حيث وضعت الترتيب بنلا من القيم المعطاة .

الجدول (١٠ - ٣)

النوع	س	ص	ف	ف'
(١)	٢	٢	٠	٠
(٢)	٤,٥	٤	٠,٥	٠,٢٥
(٣)	٩	٩	٠	٠
(٤)	٤,٥	٦	١,٥-	٢,٢٥
(٥)	٣	٣	٠	٠
(٦)	١	١	٠	٠
(٧)	٧	٨	١-	١
(٨)	٨	٧	١	١
(٩)	٦	٥	١	١
(١٠)	١٠	١٠	٠	٠
			صفر	٥,٥

$$\text{بالتصويص في (١٤) نجد أن } ١ - \frac{٥,٥ \times ٦}{٩٩ \times ١٠} = ٠,٩٦٦٧$$



#### ملاحظة (٤) :

إذا كان لمفردتين أو أكثر نفس القيمة العددية ، تعطى لكل منها رتبة تساوى الوسط الحسابي للرتب التي كانت ستأخذها هذه المفردات لو أنها كانت مختلفة القيم ، فمثلا في محتوى القار لدينا عدداً متساويان ١٧ ، ١٧ والمفروض أن أحدهما الرابع والآخر الخامس ولذلك أعطى لكل منهما الترتيب  $\frac{1}{2}(4 + 5) = 4,5$  .

#### ملاحظة (٥) :

إذا كانت البيانات أصلاً على هيئة أزواج من الترتيب كما في المثال (١٠ - ٣) فإن معامل الارتباط العزمي يساوى بالضبط معامل ارتباط الرتب . أما إذا كانت البيانات أصلاً على هيئة أزواج من القيم العددية كما في المثال (١٠ - ٤) ، فإن معامل الارتباط العزمي لها لا يساوى معامل ارتباط الرتب الناتج من تحويل هذه القيم إلى رتب .

#### (٩ - ١٠) دلالة معامل ارتباط الرتب :

إن الإحصاءة التي يعبر عنها معامل ارتباط الرتب  $r_s$  المعروف بالصيغة (١٤) توزيعها متماثل حول الصفر . ويمكن إثبات أنه إذا كان المتغيران  $x$  ،  $y$  مستقلين فإن هذا التوزيع يقترب من توزيع معتدل وسطه الحسابي صفر وتباينه  $\frac{1}{n-1}$  حين تقترب  $n$  من اللانهاية . وعلى ذلك عندما يكون حجم العينة كبيراً ( $n > 30$ ) نستطيع أن نختبر ما إذا كان هناك ارتباط ذو دلالة بين المتغيرين وذلك بحساب القيمة .

$$t_s = \frac{r_s - 0}{\sqrt{\frac{1}{n-1}}} = \frac{r_s}{\sqrt{\frac{1}{n-1}}}$$

ومقارنتها بالقيم الحرجة للتوزيع المعتدل المعيارى ، مع ملاحظة أن الفرض الصفري هو  $\rho = 0$  . والفرض الآخر هو  $\rho \neq 0$  .



فها هذا القول صحيحاً ويكون التغير في أحد المتغيرين هو فعلا سبب أو أحد أسباب التغير في الآخر كما هو الحال مثلا بين درجة الحرارة وعدد ضربات القلب ، وهناك حالات لا يصح فيها الاستنتاج حيث يكون الارتباط الذي ظهر بين المتغيرين ناشعاً عن وجود متغير ثالث أو أكثر يؤثر فيهما معاً فيحدث هذا الارتباط ، كما هو الحال مثلا بين النجاح في الرياضيات والنجاح في التاريخ حيث يؤثر في كل منهما الذكاء أو المثابرة أو البيعة المنزلية وما إلى ذلك . إن تفسير وجود الارتباط لا يكون بناء على قيمة معامل الارتباط فقط بل على ما لدينا من معلومات عن المتغيرين اللذين ندرسهما .

### تمارين (١٠)

في كل من المسائل الخمسة الآتية أوجد معامل الارتباط العزمي ( بيرسون ) من العينة المعطاة واختبر ما إذا كان المتفران - ، - مرتبطين خطياً .

(١) س : ١٥٩ ١٧٩ ١٠٠ ٤٥ ٣٨٤ ٢٣٠ ١٠٠ ٣٧٠ ٨٠ ٢٧٠ ٣٧٠ ٣١٠  
ص : ١٤,٤٠ ١٥,٢٠ ١١,٣٠ ٢,٥٠ ٢٢,٧٠ ١٤,٩٠ ١,٤١ ١٥,٨١ ٤,١٩ ١٥,٢٩ ١٧,٣٥ ٩,٥٢

حيث س وزن الخيشوم بالملجرام ، ص وزن الجسم بالجرامات لعينة حجمها ١٢ من نوع من سرطان البحر ( أبو جلمبو ) .

(٢) س : ١٧ ٢٢ ٢٠ ٢٣ ٢٤ ٢٢ ٢٠ ٢١ ٢١ ١٦  
ص : ٧ ١٢ ٩ ١٤ ١٢ ١١ ١٠ ٩ ١٠ ٨

حيث س طول نوع معين من نبات البسلة بالمليمترات ، ص الوزن بالملجرام .

(٣) س : ٥٢ ٤٨ ٥٠ ٥١ ٤٧ ٥١ ٥٢ ٥٢ ٥٣ ٤٨ ٥٠ ٥٢  
 ص : ٣٦ ٣٤ ٣٤ ٣٧ ٣٦ ٣٦ ٣٥ ٣٥ ٣٤ ٣٤ ٣٤ ٣٦  
 س : ٥٤ ٥١ ٤٨ ٥٠ ٥٠ ٤٨ ٤٨ ٤٨ ٤٩ ٥٠ ٥٠ ٥٢  
 ص : ٣٨ ٣٦ ٣٤ ٣٣ ٣٥ ٣٥ ٣٣ ٣٤ ٣٤ ٣٤ ٣٥ ٣٦

حيث س طول جسم طفل حديث الولادة ، ص طول محيط رأسه  
 بالسنتيمترات .

(٤) س : ٥٨ ٥٦ ٥٧ ٥٢ ٦٢ ٥٥ ٦٠ ٥٥ ٦١ ٥٨  
 ص : ٤١ ٤٣ ٤٢ ٤٤ ٤٤ ٤٠ ٤٢ ٤٥ ٤٥ ٤٤

حيث س ، ص هما أكبر وأصغر قطر لبيض الدجاج بالمليمترات .

(٥) س : ٤,٣ ٤,٥ ٥,٩ ٥,٦ ٦,١ ٥,٢ ٣,٨ ٢,١ ٧,٥  
 ص : ١٢٦ ١٢١ ١١٦ ١١٨ ١١٤ ١١٨ ١٣٢ ١٤١ ١٠٨

حيث س مقدار المطر في اليوم (٠,٠١ من السنتيمتر) ، ص مقدار مازال من  
 تلوث الهواء نتيجة للمطر ( ميكروجرام / متر مكعب ) . ( لتسهيل الحساب اطرح  
 ١٠٠ من قيم ص ) .

(٦) طلب من اثنين من المحكمين ترتيب ١٠ أشخاص متقدمين لوظيفة ما  
 فكانت النتيجة كما يلي . أوجد معامل الارتباط ( سيرمان ) واختبر ما إذا كان هناك  
 ارتباط موجب بين المحكمين .

المحكم الأول : ٩ ٤ ٦ ٥ ٣ ١ ١٠ ٨ ٢ ٧  
 المحكم الثاني : ٨ ٤ ٩ ٦ ١ ٢ ٧ ١٠ ٣ ٥

(٧) الجدول الآتي يبين عدد الساعات ( لعينة عشوائية من ١٠ طلاب ) التي استذكر فيها هؤلاء الطلاب لاختبار ما وعدد الدرجات التي حصلوا عليها في هذا الاختبار :

عدد ساعات الدراسة : ٨ ٥ ١١ ١٣ ١٠ ٥ ١٨ ١٥ ٢ ٨  
درجة الاختبار : ٥٦ ٤٤ ٧٩ ٧٢ ٧٠ ٥٤ ٩٤ ٨٥ ٣٣ ٦٥

حول هذه القيم إلى رتب ثم أوجد معامل ارتباط الرتب ( سيرمان ) .  
اختبر ما إذا كان هناك ارتباط موجب بين عدد ساعات الدراسة ودرجة الاختبار .

(٨) في عيتين مستقلتين حجمهما ٢٣ ، ٢٨ من أزواج القيم وجد أن معامل الارتباط العزمي بين المتغيرين ٠,٥ ، ٠,٨ ، على الترتيب . اختبر ما إذا كان هذان المعاملان مختلفين اختلافا ذا دلالة .



## الفصل الحادى عشر

### تحليل التغاير

### ANALYSIS OF COVARIANCE

(١١ - ١) التغاير :

لعله من المفيد فى مستقبل هذا الفصل أن نذكر القارئ بالمقصود حسابيا بكلمة « التغاير » التى سبق أن وردت فى عدة مناسبات فى هذا الكتاب .

إذا كانت

$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$
ص <sub>١</sub>	ص <sub>٢</sub>	ص <sub>٣</sub>	ص <sub>٤</sub>	ص <sub>٥</sub>	ص <sub>٦</sub>

،

هى  $n$  من أزواج الأعداد فإن تغاير (س، ص) يعرف بأنه متوسط مجموع حواصل ضرب انحرافات القيم السنية عن متوسطها  $\bar{s}$  وانحرافات القيم الصادية عن متوسطها  $\bar{v}$  . أى أن :

$$\text{تغاير (س، ص)} = \frac{1}{n} \sum_{i=1}^n (s_i - \bar{s})(v_i - \bar{v})$$

$$= \frac{1}{n} \left[ \sum_{i=1}^n s_i v_i - \frac{(\sum_{i=1}^n s_i)(\sum_{i=1}^n v_i)}{n} \right]$$

والمقدار الذى بين القوسين يسمى مجموع حواصل ضرب الانحرافات السينية والانحرافات الصادية عن متوسطيهما أو اختصارا مجموع حواصل الضرب وسنرمز له بالرمز  $\sigma^2$  (س ، ص) . وهذا المقدار يمكن أن يكون موجبا أو سالبا أو صفرا .

وكما فى التباين ، إذا كانت أزواج القيم (س ، ص) هى عينات عشوائية من مجتمع ذى متغيرين وأردنا تقدير التباير فى هذا المجتمع من التباير فى العينة فإننا نقسم حواصل الضرب على  $n - 1$  بدلا من  $n$  وذلك لكى يكون هذا التقدير تقديرا غير متحيز ، أى نكتب :

$$(٢) \quad \text{تباير (س ، ص)} = \frac{1}{n - 1} (\sum s^2 - \frac{(\sum s)^2}{n})$$

حيث  $\sum$  مجموع السينات ،  $\sum^2$  مجموع الصادات .

وكما فى التباين أيضا ، ونظرا لأن قيم الانحرافات عن الوسط الحسابى ( أو أى قيمة أخرى ) لا تتغير بتغير نقطة الأصل فإن قيمة التباير لا تتغير إذا جمعنا أو طرحنا عددا ثابتا من جميع السينات ، وجمعنا أو طرحنا عددا ثابتا من جميع الصادات .

#### (١١ - ٢) العلاقة بين تحليل التباير وتحليل التباين :

فى تحليل التباين بالتمودج ثابت التأثيرات للتجارب ذوات العامل الواحد يكون لدينا متغير كمى  $\sigma^2$  قسمت قيمه فى عينة ما إلى عدد من المجموعات تناظر مستويات عامل تجريب نوعى مستقل عن المتغير  $\sigma^2$  . وتهدف الدراسة إلى اختبار دلالة الفروق بين المتوسطات الصادية فى هذه الأقسام لمعرفة مدى تأثير عامل



التجريب عليها . وتتخذ البيانات المشاهدة في العينة الشكل المبين بالجدول ( ٨ - ١ ) بالبنء ( ٨ - ٤ ) .

إلا أنه في بعض هذه التجارب يكون المتغير ص واقعا تحت تأثير متغير كمى ص يسمى بالمتغير الملازم للمتغير ص *Concomitant variable or covariate* ، فمثلا قد يكون المتغير ص هو درجات الطلاب في اختبار رياضيات بعد دراسة مقرر ما فيها ويكون عامل التجريب هو طرق تدريس هذا المقرر . ونظرا لأن استيعاب الطلاب للرياضيات يعتمد على ذكائهم فإن درجاتهم في الاختبار تكون متأثرة بالذكاء ونقول حيثذ إن المتغير ص ( نسب ذكاء الطلاب ) هو متغير ملازم للمتغير ص . وإذا كان اهتمامنا هو قياس أثر اختلاف مستويات عامل التجريب ( طرق التدريس ) على المتوسطات الصادية ( درجات الرياضيات ) عن طريق تحليل التباين فإن دقة هذا القياس تستلزم استبعاد أثر المتغير ص (الذكاء ) من قيم المتغير ص قبل إجراء هذه العملية .

وفي بعض التجارب يمكن استبعاد هذا الأثر قبل عملية التجريب ، وذلك بأخذ عينات الطلاب التي تختار للتجريب بحيث تكون متكافئة في الذكاء ، وهنا نستطيع اختبار الفروق بين متوسطات الدرجات بالطريقة المعتادة لتحليل التباين على أساس أن هذه المتوسطات تكون متأثرة فقط بعامل التجريب . غير أن ظروف التجريب قد لا تتيح لنا التحكم في العينات من حيث جعلها متكافئة في المتغير الملازم ( الذكاء ) وهنا يكون لهذا المتغير أثر على المتغير ص ( درجات الرياضيات ) ولا ينبغي القيام بتحليل التباين إلا بعد استبعاد هذا الأثر . وهذا هو الدور الذى يلعبه تحليل التغاير إذ هو يتألف من شقين أولهما تعديل قيم المتغير ص لاستبعاد أثر المتغير الملازم ص وثانيهما تحليل التباين للقيم الصادية المعدلة .

ولكن كيف نصصح القيم الصادية المشاهدة لإزالة أثر هذا المتغير ؟ إذا كان لهذا الأثر وجود فعلى فإن الحدار المتغير ص على المتغير ص يكون له وجود أيضاً ويكون



## بيانات تحليل العناصر

والنموذج الذي نفترضه لتحليل التغاير هو تطوير للنموذج الذي افترضناه في تحليل التباين بالبند (٨ - ٤ - ١) ، ويتخذ الصيغة الآتية :

حيث  $\beta$  معامل الانحدار ص على س،  $s = 1, \dots, n$ ،  $q = 1, 2, \dots, k$

وهذا النموذج كما نرى يفسح مكانا لأثر الانحدار ص على س بالإضافة إلى أثر مستويات عامل التجريب وأثر الخطأ العشوائى .

فى هذا النموذج نفترض الافتراضات المعتادة فى تحليل التباين - راجع البند (٨ - ٤ - ١) - وكذلك الافتراضات المعتادة لتحليل الانحدار - راجع البندين (٩ - ١) ، (٩ - ٥) . وبالإضافة إلى ذلك نفترض فرضا أساسيا فى تحليل التباين وهو أن العلاقة الخطية بين ص- ، ص- لها نفس معامل الانحدار  $\beta$  فى جميع الأقسام أى يمكن تمثيل هذه العلاقات فى الأقسام المختلفة بخطوط متوازية ، فإذا كانت  $\beta_1$  ،  $\beta_2$  ، ... ،  $\beta_r$  هى معاملات الانحدار داخل الأقسام ، فإننا نفترض أن  $\beta_1 = \beta_2 = \dots = \beta_r = \beta$  وبذلك يمكن ضم البيانات بداخل الأقسام لتعطى تقديرا موحدا للبارامتر  $\beta$  ، وهذا التقدير هو كما نعلم  $\frac{\sum (ص- ص) \sum (ص- ص)}{\sum (ص- ص)}$  من داخل الأقسام .

إذا كتبنا النموذج (٣) على الصورة

$$\begin{aligned} \text{ص-} - \beta (\text{ص-} - \text{ص-}) &= \text{خطأ} + \text{خطأ} \\ \text{أى} \quad \text{ص-} &= \beta (\text{ص-} - \text{ص-}) + \text{خطأ} + \text{خطأ} \end{aligned} \quad (٤)$$

يتبين لنا أن تحليل التباين يتطلب إجراء نوعين من التحليل هما :

١ - تحليل انحدار لتقدير قيمة  $\beta$  ومن ثم تعديل القيم ص- المشاهدة إلى القيم ص- لإزالة أثر المتغير الملائم ص- .

٢ - تحليل التباين للقيم المعدلة ص- لاختبار أثر مستويات التجريب على متوسطاتها . أى أن تحليل التباين يجمع بين عملية تحليل الانحدار وعملية تحليلها فى عملية تحليل التباين .

#### (١١ - ٤) خطوات تحليل التغيرات :

إن الهدف من تحليل التغيرات هو كما سبق القول اختبار دلالة الفروق بين متوسطات المتغير التابع  $\bar{y}$  بعد تصحيحها لاستبعاد أثر المتغير الملزم  $\bar{y}$  . ومن حقنا أن نتساءل بداية عما إذا كان للمتغير  $\bar{y}$  أثر فعل على المتغير  $\bar{y}$  بحيث يستحق عناء عملية الاستبعاد . ولذلك فإن أول ما نهتم به اختبار وجود هذه العلاقة لا يكون اختبار وجود علاقة بين  $\bar{y}$  ،  $\bar{y}$  ، فإذا لم يثبت وجود هذه العلاقة لا يكون هناك ما يدعو لعملية الاستبعاد بل نقوم بتحليل التباين على القيم الصادية المشاهدة دون تعديل . أما إذا ثبت وجودها فينبغي تصحيح هذه القيم قبل إجراء عملية تحليل التباين .

فالخطوة الأولى لتحليل التغيرات إذن هي تحليل الانحدار لاختبار أثر المتغير  $\bar{y}$  على المتغير  $\bar{y}$  ، فإذا ثبتت دلالة الانحدار فإن الخطوة الثانية هي القيام بتحليل التباين للقيم الصادية المعدلة لاختبار أثر عامل التجريب . والمعتاد أن نعهد لهاتين الخطوتين ببعض الحسابات الأولية . وسنوضح ذلك كله بالمثال الآتي .

#### المثال (١١ - ١) :

في إحدى التجارب الزراعية الاقتصادية في قطر ما ، اختيرت ست قرى عشوائية واختير في كل منها خمس مزارع عشوائية . سجلت بالجدول (١١ - ٢) الآتي تكاليف إنتاج محصول الذرة (ص) في فصل زراعي ما في كل مزرعة كما سجلت متوسطات المحاصيل (س) التي كانت قد نتجت في فصول زراعية سابقة في كل مزرعة . المطلوب استخدام هذه البيانات (أولا) لمعرفة ما إذا كان هناك أثر للمحصول السابق (س) على تكاليف الإنتاج (ص) ، (ثانيا) لمعرفة ما إذا كانت تكاليف الإنتاج تختلف من قرية إلى أخرى بعد استبعاد أثر مقدار المحصول السابق إذا كان لهذا الأثر وجود . ( لتسهيل الحساب طُرح العدد ١٢ من جميع قيم س والعدد ٣ من جميع قيم ص ) .

المجمول (١١ - ٢)

القرى المزارع	(١)	(٢)	(٣)	(٤)	(٥)	(٦)
	ص <sub>١</sub> ص <sub>٢</sub>	ص <sub>١</sub> ص <sub>٢</sub>	ص <sub>١</sub> ص <sub>٢</sub>	ص <sub>١</sub> ص <sub>٢</sub>	ص <sub>١</sub> ص <sub>٢</sub>	ص <sub>١</sub> ص <sub>٢</sub>
(١)	١,٧ ٣	٠,٨ ٠	١,٥ ٣-	٠,٢- ٤	١,٧ ٢-	٠,٥ ١
(٢)	٢,١ ١-	٠,٨- ٥	٠,٦ ٠	١	٢,٣ ٣-	١ ١-
(٣)	٣,٠ ٣-	١,٠- ٧	١,٠ ٢-	٠,٢- ٣	٢,٦ ٤-	٠,٥- ٦
(٤)	١,٣ ٢	١,٦ ١-	٠,٤ ١	١,٢ ٢-	١,٠	٠,٧ ٢
(٥)	٠,٩ ٥	٠,٤ ٢	٢,٥ ٤-	٠,٢ ٠	١,٤ ١-	١,٣ ٠
المجموع	٩ ٦	١ ١٣	٦ ٨-	١ ٦	٩ ١٠-	٣ ٨
التوسطات	١,٨ ١,٢	٠,٢ ٢,٦	١,٢ ١,٦-	٠,٢ ١,٢	١,٨ ٢-	٠,٦ ١,٦

$\bar{ص} = ١٥$  ،  $\bar{ص} = ٠,٥$  ،  $\bar{ص} = ٢٩$  ،  $\bar{ص} = ٠,٩٦٦٧$  ،  
 مع ملاحظة أن  $٣٠ = \bar{ن}$

حسابات قهيدية :

هذه الحسابات تجرى فى بداية التحليل لخدمة الخطوتين الرئيسيتين المذكورتين ،  
 وهى تتألف من ثلاث مجموعات من الحسابات ترمى المجموعة الأولى إلى إيجاد  
 مجموع المربعات  $\sum (ص)$  للمتغير  $ص$  وسنرمز له بالرمز  $\sum$  ثم تحليله بالطريقة المعتادة  
 لتحليل التباين إلى مركبتين إحداهما تعبر عن الاختلاف بين الأقسام وسنرمز له  
 بالرمز  $\sum$  ، والأخرى تعبر عن الاختلاف المتبقى داخل الأقسام وسنرمز له بالرمز



(٢) تحليل ٢ ٢ (ص)

$$١ - ١ = ٢ = ٢ (الكلي) = ٢ - ٢ = \frac{٢}{٢} - \frac{٢}{٢} \quad \text{بدرجات حرية } ١ - ١$$

$$١ - ١ = ٢ = ٢ (بين الأقسام) = \frac{٢}{١} - \frac{٢}{١} = \frac{٢}{١} - \frac{٢}{١} \quad \text{بدرجات حرية } ١ - ١$$

حيث تم مجموع الصادات في القسم ١ ، ١ ، عددها .

$$١ - ١ = ٢ = ٢ (داخل الأقسام) = ١ - ١ = ١ - ١ \quad \text{بدرجات حرية } ١ - ١$$

(٣) تحليل ٢ ٢ (س ، ص)

تسير العمليات الجبرية هنا في خطوط متوازية مع العمليات الجبرية في (١) ،  
(٢) كالآتي :

$$١ - ١ = ٢ = ٢ (الكلي) = ٢ - ٢ = \frac{٢}{٢} - \frac{٢}{٢} \quad \text{بدرجات حرية } ١ - ١$$

$$١ - ١ = ٢ = ٢ (بين الأقسام) = \frac{٢}{١} - \frac{٢}{١} = \frac{٢}{١} - \frac{٢}{١} \quad \text{بدرجات حرية } ١ - ١$$

$$١ - ١ = ٢ = ٢ (داخل الأقسام) = ١ - ١ = ١ - ١ \quad \text{بدرجات حرية } ١ - ١$$

بتطبيق هذه الصيغ على بيانات المثال (١١ - ١) نحصل على مايلي :

(١) تحليل ٢ ٢ (س) :

$$١٥ = ٨ + (١٠-) + ٦ + (٨-) + ١٣ + ٦ = ٢ \quad \text{لدينا}$$

$$\therefore \frac{١٥}{٣٠} = \frac{٢}{١٠}$$



$$٧,٥ - [٠ + '٢) + ... + '١-) + '٣] = \text{٢٢ (الكل)}$$

$$\text{بلرجات حرية ٢٩} \quad ٢٥١,٥ =$$

$$+ '٦ + '٨-) + '١٣ + '٦] \frac{1}{٥} = \text{٢٢ = (بين الأقسام)}$$

$$٧,٥ - [٨ + '١٠-)$$

$$\text{بلرجات حرية ٥} \quad ٨٦,٣ =$$

$$\text{٢٤ = (داخل الأقسام)} \quad ٨٦,٣ - ٢٥١,٥ = ١٦٥,٢ \text{ بلرجات حرية ٢٤}$$

(٢) تحليل ٢٢ (ص) :

$$\text{لدينا } \bar{m} = ٣ + ٩ + ١ + ٦ + ١ + ٩ = ٢٩$$

$$\therefore \frac{\bar{m}}{٣٠} = \frac{٢٩}{٣٠} = ٢٨,٠٣$$

$$\text{ب = ٢٢ (الكل)} \quad [١,٣ + '١٠,٧) + ... + '٢,١ + '١,٧] =$$

$$٢٨,٠٣ -$$

$$\text{بلرجات حرية ٢٩} \quad ٢٨,٩٣ =$$

$$\text{ب = ٢٢ (بين الأقسام)} \quad [٣ + '٩ + '١ + '٦ + '١ + '٩] \frac{1}{٥} =$$

$$٢٨,٠٣ -$$

$$\text{بلرجات حرية ٥} \quad ١٣,٧٧ =$$

$$\text{ب = ٢٢ (داخل الأقسام)} \quad ١٣,٧٧ - ٢٨,٩٣ = ١٥,١٦$$

$$\text{بلرجات حرية ٢٤}$$

(٣) تحليل ٢ ص (س ، ص) :

$$١٤,٥ = \frac{٢٩ \times ١٥}{٣٠} = \frac{٢}{٣} \text{ واذن } ٢٩ = \bar{م} , ١٥ = \bar{ن}$$

$$\times ٢ + \dots + (٢,١ \times ١-) + ١,٧ \times ٣] = \text{ح} = ٢ \text{ ص (الكل)}$$

$$١٤,٥ - [٠ + ٠,٧$$

$$٦٩,٣- = ١٤,٥ - ٥٤,٨ - =$$

بدرجات حرية ٢٩

$$\text{ح}_١ = ٢ \text{ ص (بين الأقسام)} = \frac{[٣ \times ٨ + \dots ١ \times ١٣ + ٩ \times ٦]}{٥} =$$

$$١٤,٥ -$$

$$٢٢,٧- = ١٤,٥ - ٨,٢ - =$$

$$\text{ح}_٢ = ٢ \text{ ص (داخل الأقسام)} = ٦٩,٣- - (٢٢,٧-) = ٤٦,٦ =$$

بدرجات حرية ٢٤

يمكن تسجيل هذه المجموعات الثلاث من القيم في جدول واحد كالجدول (١١ - ٣) الآتي ليسهل الرجوع إليها .

الجدول (١١ - ٣)

تحليل مجموعي للربعات ومجموع حواصل الضرب

مصدر التباين	درجات الحرية	٢٢ (س)	٢٢ (ص)	٢ ص (س ، ص)
بين الأقسام (القرى)	٥ - ١ =	٨٦,٣ = ١	١٣,٧٧ = ١	٢٢,٧- = ١
داخل الأقسام (البواقي) خطأ التجريب	٢٤ = ٥ -	١٦٥,٢ = ١	١٥,١٦ = ٢	٤٦,٦- = ٢
الكل	٢٩ = ١ -	٢٥١,٥ = ١	٢٨,٩٣ = ٣	٦٩,٣- = ٣

الخطوة الأولى : اختبار تأثير المتغير م على المتغير ه :

تهدف هذه الخطوة إلى اختبار وجود علاقة خطية بين المتغير الملائم م والمتغير التابع ه ، ويتأتى هذا كما نعلم إما بتقدير معامل الانحدار الموحد  $\beta$  من داخل أقسام العينة ثم اختبار دلالة هذا التقدير باختبار ت ، أو بتقدير التباين المفسر ( الناشئ عن الانحدار ) واختبار دلالة اختبار ف بالصورة

$$F = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} \quad \text{بدرجتي حرية ١ ، ه - ك - ١ (٥)}$$

على أن تحسب المقادير اللازمة لهذا الاختبار من داخل الأقسام أى من السطر الخاص بالبوابات بالجدول (١١ - ٣) .

في المثال ، وباستخدام الصيغة (١٧) بالبند (٩ - ٧) نجد ما يلي :

$$\frac{F}{1} = \frac{[ (ص ، ه) (ص ، ه) ]}{(ص) ٢ ٢} = \text{الاختلاف المفسر}$$

$$13,145 = \frac{(46,7-)}{165,2} = \text{بدرجة حرية واحدة}$$

∴ الاختلاف غير المفسر = الاختلاف في ص - الاختلاف المفسر

$$13,145 - 15,16 = \frac{F}{1} - F =$$

$$2,015 = \text{بدرجات حرية ٢٣}$$

$$F^{**} 150,06 = \frac{13,145}{0,0876} = \frac{1 \div 13,145}{23 \div 2,015} = F \quad \text{إذن}$$

وهذه القيمة تزيد كثيرا عن القيمة الحرجة ف (١٠,٠٠١، ٢٣) فهي ذات دلالة عالية ونحكم بوجود علاقة خطية سالبة بين المتغير الملزم  $\bar{m}$  والمتغير التابع  $\bar{y}$  ، وهذا يعنى أن مقدار المحصول السابق يؤثر فيما تدفعه القرى من تكاليف لإنتاج المحصول الجديد .

#### ملاحظات :

(١) استخدمنا نفس الصيغ التي تستخدم في حالة وجود عينة من أزواج القيم ( $\bar{y}$  ،  $\bar{x}$ ) موضوعة في قسم واحد ( $k = 1$ ) لأن هذه الصيغ تظل صالحة عند وجود أكثر من قسم ( $k < 1$ ) ، بشرط تعديل درجات الحرية وفقا لذلك أى وضع  $n - k - 1$  بدلا من  $n - 2$  للاختلاف غير المفسر .

(٢) من الجدول (١١ - ٣) نستطيع إيجاد ثلاثة تقديرات لتباين المتغير  $\bar{m}$  بعد التصحيح للانحدار ، غير أن التقدير الذى نحصل عليه من داخل الأقسام وهو تباين البواقي حول خط الانحدار يكون هو التقدير غير التحيز الذى يعكس الخطأ العشوائى في هذا التحليل ، ونختار دلالة أى تقدير آخر بالمقارنة به . ولذلك فإن حساب قيمة  $F$  من الصيغة (٥) ينبغي أن يكون من داخل الأقسام كما سبق القول .

#### الخطوة الثانية : تعديل المتوسطات واختبار دلالة الفروق بينها :

كما سبق القول ، لا نتخذ هذه الخطوة إلا إذا ظهر من الخطوة السابقة وجود أثر فعلى للمتغير الملزم  $\bar{m}$  على المتغير  $\bar{y}$  . وترمى هذه الخطوة إلى تعديل القيم الصادية المشاهدة بحيث تكون القيم المعدلة مستقلة عن أثر المتغير  $\bar{m}$  . ولما كان هذا التعديل من شأنه تعديل تباين القيم الصادية فإن تحليل التباين لاختبار دلالة الفروق بين المتوسطات المعدلة يكون باستخدام اختبار  $F$  بالصيغة المعتادة مع وضع التباينات المعدلة بدلا من التباينات الأصلية ، أى باستخدام الصيغة

$$F = \frac{\text{التباين المعدل بين الأقسام}}{\text{التباين المعدل داخل الأقسام}} \quad (٦)$$

بدرجتي حرية ك - ١ ، ١ - ١ - ك - ١

في المثال (١١ - ١) وجدنا أن المتغير الملازم من يؤثر في المتغير التابع من وعلى ذلك فإن دقة البحث تقتضي أن نخلص المتغير من أثر المتغير من قبل إجراء تحليل التباين ، أي تقتضي استخدام اختبار ف بالصيغة (٦) . ولقد سبق لنا حساب التباين المعدل داخل الأقسام وهو التباين غير المفسر  $2,015 \div 23 = 0,0876$  وهو كما سبق القول تقدير غير متحيز لتباين المتغير من بعد تعديله .

ومن الصيغة (١٩) بالهند (٩ - ٧) نذكر أن هذا التباين هو مربع الخطأ المعياري ع<sub>١</sub> للتقدير من معادلة الانحدار ، أي أن :

$$\begin{aligned} \text{ع}^2_{١} &= \text{التباين غير المفسر ( مأخوذا من داخل الأقسام )} \\ (٧) \quad 0,0876 &= \text{في هذا المثال. وبدرجات حرية } 23 \end{aligned}$$

أما التباين المعدل بين الأقسام فنوجده بطريقة غير مباشرة بطرح الاختلاف المعدل داخل الأقسام من الاختلاف الكلي المعدل ثم القسمة على درجات الحرية . أي أن :

(٨) داخل الأقسام

$$(٨) \quad = (ب - \frac{\sum}{1}) - (ب - \frac{\sum}{1})$$

وفي هذا المثال نجد أن

$$2,015 - ( \frac{(19,3-)}{251,5} - 28,93) = \text{الاختلاف المعدل بين الأقسام}$$

$$2,015 - - (19,095 - 28,93) =$$

$$7,820 = \text{بدرجات حرية } 5$$

$$\frac{1,064}{0,0876} = \frac{5 \div 7,820}{23 \div 2,010} = \text{من الصيغة (٦) : في}$$

$$17,85 =$$

وهذه القيمة ذات دلالة عالية لأن ف (٢٣,٠٥) لا تزيد عن ٣,٩ مما يشير إلى أن متوسط تكاليف الإنتاج - بعد استبعاد أثر مقادير المحاصيل السابقة - ليست متساوية في القرى الست . ويمكننا إذا أردنا أن نضع هذه النتائج في الجدول (١١ - ٤) الآتي .

الجدول (١١ - ٤)

أصناف للظاير

مصدر الاختلاف	الاختلاف للشاهد ب	الانحدار ب/١	الاختلاف المعدل ب- ب/١	د ح	ط ز	ف
بين الأقسام (القرى)	-	-	٧,٨٢٠	٥	١,٥٦٤	١٧,٨٥
داخل الأقسام	١٥,١٦	١٣,١٤٥	٢,٠١٥	٢٣	٠,٠٨٧٦	
الكل	٢٨,٩٣	١٩,٠٩٥	٩,٨٣٥	٢٨		

ملاحظة :

إذا أهملنا المتغير الملزم - وقمنا بتحليل التباين لقيم - المشاهدة دون تعديل نجد من بيانات الجدول (١١ - ٣) أن

$$\frac{2,754}{0,6317} = \frac{5 \div 13,77}{24 \div 15,16} = \frac{\text{التباين بين الأقسام}}{\text{التباين داخل الأقسام}} = \text{ف}$$

$$= 4,36^{**} \quad \text{بدرجتي حرية 5 ، 24}$$

وبالرغم من أن هذه القيمة ذات دلالة عالية أيضا لأن  $F_{[24, 5], 0.01} = 3,90$  إلا أنها تقل كثيرا عن القيمة 17,85. هذا مع تذكر أن تحليل التباين يجزئ الاختلاف الكلي في ص إلى المركبتين بين وداخل الأقسام ، أما تحليل التباين فهو يجزئ البواقي من التحليل الشامل للانحدار .

#### (١١ - ٥) المقارنة بين المتوسطات المعدلة :

بعد تحليل التباين للقيم الصادية المعدلة يبقى أن نتدارس كيفية إجراء المقارنات بين متوسطات هذه القيم في الأقسام المختلفة ، وهذا يتطلب أن نحسب هذه المتوسطات ثم نختبر دلالة الفروق بينها .

#### (١) حساب المتوسطات المعدلة :

من الصيغة (٢) بالبند (١١ - ٢) نجد أن المتوسط المعدل  $\bar{ص}$  في أي قسم يساوي المتوسط المشاهد المناظر  $\bar{ص}$  مطروحا منه أثر الانحدار الخطي . أي أن

$$\bar{ص} = \bar{ص} - \bar{ب} (\bar{ص} - \bar{ص}) \quad (٩)$$

حيث  $\bar{ب}$  معامل انحدار  $\bar{ص}$  على  $\bar{ص}$  ويحسب من داخل الأقسام بالصيغة المعتادة للانحدار الخطي البسيط كالآتي - انظر الصيغة (٨) بالبند (٩ - ٣) .

$$\bar{ب} = \frac{\sum_{i=1}^r \bar{ص}_i (\bar{ص}_i - \bar{ص})}{\sum_{i=1}^r (\bar{ص}_i - \bar{ص})^2} \quad (١٠)$$

$$\text{ففي المثال : } \bar{ب} = \frac{46,6-}{165,2} = -0,282$$

من جدول البيانات (١١ - ٢) والصيغة (٩) نجد أن المتوسطات المعدلة هي :

$$\bar{ص}_1 = ١,٨ + ٠,٢٨٢ (١,٢ - ٠,٥) = ١,٩٩٧٤$$

$$\bar{ص}_٢ = ٠,٢ + ٠,٢٨٢ (٢,٦ - ٠,٥) = ٠,٧٩٢٢$$

$$\bar{ص}_٣ = ١,٢ + ٠,٢٨٢ (-١,٦ - ٠,٥) = ٠,٦٠٧٨$$

$$\text{وبالمثل نجد أن } \bar{ص}_٤ = ٠,٣٩٧٤ ، \bar{ص}_٥ = ١,٠٩٥ ، \bar{ص}_٦ = ٠,٩١٠٢$$

يلاحظ أن المجموع الكلي للقيم الصادية المعدلة يساوي المجموع الكلي للقيم الصادية المشاهدة وكلاهما يساوي ٢٩ . كما يلاحظ أن بعض المتوسطات صغرت والبعض الآخر كبر ، وأن الفرق بين أى متوسطين معدلين يقل عن الفرق بين المتوسطين المشاهدين المناظرين ، مما يشير إلى أن المتغير الملازم كان له تأثير فعلى على هذه المتوسطات .

(ب) اختبار دلالة الفروق بين أزواج المتوسطات المعدلة :

يلزمنا هنا إيجاد تقدير للخطأ المعياري  $\sigma_c$  للفرق بين متوسطين معدلين ، بحيث ندخل في اعتبارنا الخطأ المعياري  $\sigma_{مر}$  للتقدير من معادلة الانحدار . والصيغة التي نشق منها الخطأ المعياري المطلوب هي :

$$\sigma_c = \sigma_{مر} (1 + \frac{1}{r} \frac{(k-1)}{2}) \text{ بدرجات حرية } n - k - 1 \text{ (١١)}$$

حيث  $r$  = مجموع المربعات لقيم  $s$  محسوبا من بين أقسام المعالجة ،  
 $r$  = مجموع المربعات لقيم  $s$  محسوبا من داخل أقسام المعالجة ،  
 ففى المثال ، وباستخدام الجدول (١١ - ٣) نجد أن

$$\sigma_c = ٠,٠٨٧٦ (1 + \frac{1}{160,2} \frac{٨٦,٣}{٥}) = ٠,٠٩٦٧٥ \text{ بدرجات حرية } ٢٣$$



ونكون الآن مستعدين لإجراء المقارنات بين أزواج المتوسطات الصادية المعدلة باستخدام الأسلوب المبين بالبند (٨ - ٥) أو أى أسلوب مكافئ .

فمثلا للمقارنة بين المتوسطين المعدلين  $\bar{ص}_1 = ١,٩٩٧٤$  ،  $\bar{ص}_٢ = ٠,٧٩٢٢$  مع ملاحظة أن مجموعى القيم الصادية فى القسمين (١) ، (٢) هما ( بالضرب فى خمسة ) ٩,٩٨٧ ، ٣,٩٦١ يمكن أن نستخدم الأسلوب الآتى :

$$٢٢ (١ ضد ٢) = \frac{٣,٩٦١ + ٩,٩٨٧}{٥} - \frac{١٣,٩٤٥}{١٠} = ٣,٦٢٩٣$$

$$\therefore \text{فى } ٣٧,٥١٢ = \frac{١ \div ٣,٦٢٩٣}{٠,٠٩٦٧٥} \text{ بدرجتى حرية ١ ، ٢٣ .}$$

وهذه القيمة ذات دلالة عالية مما يجعلنا نرفض الفرض الصفرى عن تساوى متوسطى التكاليف فى القريتين (١) ، (٢) .

كما يمكننا هنا استخدام اختبار ت بالصيغة

$$ت = \frac{\bar{ص}_١ - \bar{ص}_٢}{\sqrt{\frac{١}{١٠} + \frac{١}{١٠}}} \text{ بدرجات حرية ٢٣}$$

فنجد أن  $ت = ٦,١٢٧$  وهى أيضا ذات دلالة عالية ، مع ملاحظة أن مربع هذه القيمة وهو ٣٧,٥٤ يناوى قيمة  $ف = ٣٧,٥١٢$  والفرق يرجع إلى أخطاء التقريب .

## تقريب (١١ - ١)

فى جزء من تجربة عن تأثير نوعين من الأدوية فى علاج مرض الجذام كان الهدف مقارنة ثلاثة أنواع من المعالجة : د ، د١ ، د٢ دواءان من صنف المضادات الحيوية ،

د. دواء داخلي يتخذ كمراقبة . اختبر عشرة مرضى للتجريب وحددت على كل منهم ٦ مواقع من الجسم يتراكم فيها ميكروب الجذام وقيست غزارة الجذام باختبارات معملية في هذه المواقع قبل بداية التجربة (س) ثم بعد عدة شهور من العلاج (ص) ودونت النتائج في الجدول الآتي .

الجدول (١١ - ٥)

المرضى	الدواء					
	د <sub>١</sub> ص <sub>١</sub>	د <sub>٢</sub> ص <sub>٢</sub>	د <sub>٣</sub> ص <sub>٣</sub>	د <sub>٤</sub> ص <sub>٤</sub>	د <sub>٥</sub> ص <sub>٥</sub>	د <sub>٦</sub> ص <sub>٦</sub>
(١)	١١	٦	٠	٦	١٦	١٣
(٢)	٨	٠	٦	٢	١٣	١٠
(٣)	٥	٢	٧	٣	١١	١٨
(٤)	١٤	٨	٨	١	٩	٥
(٥)	١٩	١١	١٨	١٨	٢١	٢٣
(٦)	٦	٤	٨	٤	١٦	١٢
(٧)	١٠	١٣	١٩	١٨	١٢	٥
(٨)	٦	١	٨	٩	١٢	١٦
(٩)	١١	٨	٥	١	٧	١
(١٠)	٣	٠	١٥	٩	١٢	٢٠
المجموع	٩٣	٥٣	١٠٠	٦١	١٢٩	١٢٣
المتوسطات	٩,٣	٥,٣	١٠	٦,١	١٢,٩	١٢,٣

أجر تحليل التباين لاختبار دلالة الفروق بين متوسطات غزارة الجذام في المستويات الثلاثة للمعالجة ، بعد استبعاد أثر الاختلاف في غزارة المرض قبل بدء المعالجة .

(١١ - ٦) تحليل التباين في التجارب ذوات العاملين ومتغير ملازم واحد :

النموذج الإحصائي هنا هو امتداد طبيعي للنموذج الإحصائي (٣) للتجارب ذوات العامل الواحد ، وهو يأخذ الصيغة الآتية :

$$ص_{ري} = \mu + \alpha + \beta(س_{ري} - \bar{س}) + \gamma \quad (١٢)$$

ويسير التحليل بنفس الأسلوب مع بعض الإضافات والتعديلات التي يقتضيها وجود عامل تجريبي ثان .

مثال (١١ - ٢) :

في تجربة لمقارنة مقادير المحاصيل الناتجة من ٦ أنواع من الذرة أخذت عينة عشوائية من ٢٤ حوضاً زراعياً وزرعت عليها أنواع الذرة عشوائياً - أربعة أحواض لكل نوع - وقد نتجت البيانات المدونة بالجدول (١١ - ٦) الآتي حيث ص تعبر عن مقدار المحصول الناتج من الحوض ، س تعبر عن درجة خصوبة الحوض مقدرة بعدد النباتات التي كانت قائمة به . إن مقدار المحصول يكون واقعاً تحت تأثير عاملين تجريبيين هما : الأحواض (٤ مستويات) وأنواع الذرة (٦ مستويات) ، غير أن المطلوب هنا اختبار ما إذا كان معدل المحصول يختلف باختلاف نوع الذرة (بعد استبعاد أثر الخصوبة) .

الحسابات التمهيدية :

نقوم بتحليل كل من ٢٢ (س) ، ٢٢ (ص) ، ٢٢ (ص س) ، إلى ثلاث مركبات : بين الأعمدة وبين الصفوف والخطأ . أي أن هناك خطوة إضافية واحدة في كل تحليل .

المجلد (١١-٦)

أنواع الفترة	الأحواض				الجميع	المتوسطات $\bar{x}$ $\bar{s}$
	(١) س. ص.	(٢) س. ص.	(٣) س. ص.	(٤) س. ص.		
أ	٢٠٢ ٢٨	١٦٥ ٢٢	١٩١ ٢٧	١٣٤ ١٩	٦٩٢ ٩٦	١٧٣ ٢٤
ب	١٤٥ ٢٣	٢٠١ ٢٦	٢٠٣ ٢٨	١٨٠ ٢٤	٧٢٩ ١٠١	١٨٢,٢٥ ٢٥,٥
ج	١٨٨ ٢٧	١٨٥ ٢٤	١٨٥ ٢٧	٢٢٠ ٢٨	٧٧٨ ١٠٦	١٩٤,٥ ٢٦,٥
د	٢١٠ ٢٤	٢٣١ ٢٨	٢٣٨ ٣٠	٢٦١ ٣٠	٩٣١ ١١٢	٢٣٢,٧٥ ٢٨
هـ	٢٠٢ ٣٠	١٧٨ ٢٦	١٩٨ ٢٦	٢٢٦ ٢٩	٨٠٤ ١١١	٢٠١ ٢٧,٧٥
و	٢٢٨ ٣٠	٢٢١ ٢٥	٢٠٧ ٢٧	٢٠٤ ٢٤	٨٦٠ ١٠٦	٢١٥ ٢٦,٥
الجميع	١١٦٦ ١٦٢	١١٨١ ١٥١	١٢٢٢ ١٦٥	١٢٢٥ ١٥٤	٤٧٩٤ ٦٣٢	١٩٩,٧٥ ٢٦,٣٢

م.ح.س = ١٦٨٢٤ ، م.ح.ص = ٩٧٦٢٨٠ ، م.ح.س.ص = ١٦٧٧٢٧

تحليل ٢٢ (س) :

$$١ = ٢٢ (الكُل) = م.ح.س - \frac{م.ح.س.ص}{٢٤}$$

$$٢٣ = ١ - \text{بدرجات حرية } ٢٤ - ١ = ٢٣ \quad ١٨١,٣٣٤ = \frac{٦٣٢}{٢٤} - ١٦٨٢٤ =$$

$$١ = ٢٢ (بين الأعمدة) = \frac{١}{٦} (١٦٢ + ١٥١ + ١٦٥ + ١٥٤) - \frac{٦٣٢}{٢٤}$$

$$٣ = ١ - \text{بدرجات حرية } ٢٤ - ٢١ = ٣$$

$$\frac{1}{4} = \frac{1}{4} (٢٢ \text{ بين الصفوف}) = \frac{1}{4} (٩٦ + ١٠١ + \dots + ١٠٦) - \frac{٦٣٢}{٢٤}$$

$$٤٥,٨٣٤ = \text{بلرجات حرية ه} - ١ = ٥$$

$$\therefore ٢٢ = \text{! (الخطأ)} = ١٨١,٣٣٤ - (٢١,٦٦٧ + ٤٥,٨٣٤)$$

$$١٥ = (١ - ه) (١ - ك) \text{ بلرجات حرية (ك) (هـ) (١ - هـ) = ١٥}$$

وبنفس الطريقة نحل كلا من ٢٢ (ص) و ٢٣ (س، ص) ونضع النتائج في الجدول الآتي حيث ك ترمز إلى عدد الأعمدة ، ه ترمز إلى عدد الصفوف .

الجدول (١١-٧)

تحليل مجموعي المربعات ومجموع حواصل الضرب

مصدر التباين	درجات الحرية	٢٢ (س)	٢٢ (ص)	٢٣ (س، ص)
بين الأعمدة (الأحادي)	٣ - ١ - ٥	٢١,٦٧ = أ	٤٣٦,١٧ = ب	٨,٥٠ = ج
بين الصفوف (أنواع)	٥ - ١ - ٥	٤٥,٨٣ = أ	٩٤٩٠,٠٠ = ب	٥٥٩,٢٥ = ج
خطأ التجريب	١٥ = (١-هـ)(١-ك)	١١٣,٨٣ = أ	٨٧٥٢,٣٣ = ب	٩١٧,٢٥ = ج
أنواع + خطأ	٢٠	١٥٩,٦٦	١٨٢٤٢,٣٣	١٤٧٦,٥
الكل	٢٣ - ١ - ٥	١٨١,٣٣ = أ	١٨٦٧٨,٥٠ = ب	١٤٨٥,٠٠ = ج

وستبين أهمية السطر الإضافي (أنواع + خطأ) عند تكوين النسبة ف لاختبار المتوسطات المعدلة .

الخطوة الأولى : اختبار تأثير المتغير اللازم  $\mu$  على المتغير  $\nu$  :

أى اختبار وجود علاقة خطية بين المتغيرين  $\mu$  ،  $\nu$  ويتأتى هذا عن طريق الصيغة (٥) لاختبار ف مع ملاحظة درجات الحرية .

$$F = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} \quad \text{بلرئى حرية ١ ، ٧ - ٤ - ٥ (١٣)}$$

على أن يحسب البسط والمقام من السطر الخاص بخطأ التجريب . فى المثال نجد ما يأتى :

$$\text{الاختلاف المفسر} = \frac{917,25}{113,83} = \frac{8}{1} = 8$$

بلرئة حرية واحدة

$$\therefore \text{الاختلاف غير المفسر} = 8 - 8 = 0$$

$$= 1361,0662 \quad \text{بلرئى حرية ١٤}$$

$$\therefore F = \frac{8}{0} = \frac{1 \div 8}{14 \div 1361,0662} = \frac{1}{10087,2564}$$

$$= 76,027 \quad \text{بلرئى حرية ١ ، ١٤}$$

وهذه القيمة ذات دلالة عالية مما يدل على وجود تأثير كبير لخصوبة الأرض على معدل المحصول . ولذلك ينبغي استبعاد هذا الأثر قبل مقارنة متوسطات المحاصيل تحت تأثير أنواع النرة .

الخطوة الثانية : اختبار دلالة الفروق بين المتوسطات المعدلة :

نستخدم الصيغة (٦) لاختبار ف مع ملاحظة درجات الحرية .

$$ف = \frac{\text{التباين المعدل (بين الأنواع)}}{\text{التباين المعدل للخطأ}} \text{ بـدرجتي حرية ك-١ ، د-ك هـ (١٤)}$$

ولقد سبق أن حسبنا في الخطوة الأولى التباين المعدل للخطأ فهو التباين غير المفسر

$$٩٧,٢١٩ = ١٤ \div ١٣٦١,٠٦٦٢$$

أما التباين المعدل بين الأنواع فيحسب كالآتي :

التباين المعدل بين الأنواع = [ الاختلاف المعدل (أنواع + خطأ) - الاختلاف

المعدل للخطأ ] مقسوما على درجات الحرية (١٥)

من السطر الإضافي بالجدول (١١ - ٧) نجد ما يلي

$$\frac{١٤٧٦,٥}{١٥٩,٦٦} - ١٨٢٤٢,٣٣ = (\text{أنواع} + \text{خطأ}) \text{ الاختلاف المعدل}$$

$$٤٥٨٧,٩٨٩ = \text{بـدرجات حرية ١٩}$$

$$١٣٦١,٠٦٦٢ = \text{بـدرجات حرية ١٤} \text{ ولكن الاختلاف المعدل للخطأ}$$

$$١٣٦١,٠٦٦٢ - ٤٥٨٧,٩٨٩ = \text{الاختلاف المعدل بين الأنواع}$$

$$٣٢٢٦,٩٢٢٨ = \text{بـدرجات حرية ٥}$$

$$٦٤٥,٣٨٥ = ٥ \div ٣٢٢٦,٩٢٢٨ = \text{التباين المعدل بين الأنواع}$$

من الصيغة (١٤) ينتج أن :

$$ف = \frac{٦٤٥,٣٨٥}{٩٧,٢١٩} = ٦,٦٣٨ \text{ بـدرجتي حرية ٥ ، ١٤}$$

وهذه القيمة ذات دلالة عالية لأنها أكبر من ٠.٠٥ ، ٠.٠١ ، ٠.٠٠١ التي لا تزيد عن ٥,٠٦ مما يشير إلى أن متوسطات المحاصيل ( بعد استبعاد أثر الخصوبة ) ليست جميعها متساوية عند الأنواع المختلفة من الذرة .  
ملاحظة :

يمكن بنفس الطريقة اختبار ما إذا كان معدل المحصول يختلف باختلاف الأحواض ( الأعمدة ) .

( ١١ - ٧ ) المقارنة بين أزواج المتوسطات المعدلة :

نستخدم نفس الأسلوب المقدم بالبند ( ١١ - ٥ - ب ) ، فتحسب المتوسطات المعدلة من الصيغة (٩) وهي :

$$\overline{ص_1} = \overline{ص_2} - \overline{ص_3} - \overline{ص_4} + \overline{ص_5} = \overline{ص_1} - \overline{ص_2} + \overline{ص_3} - \overline{ص_4} + \overline{ص_5}$$

$$٨,٠٥٨٨ = \frac{٩١٧,٢٥}{١١٣,٨٢} = \frac{١}{١} = \frac{٢}{٢} = \frac{٣}{٣} = \frac{٤}{٤} = \frac{٥}{٥}$$

ومن هذه القيمة نتوصل إلى المتوسطات المعدلة الآتية :

$$\overline{ص_1} = ١٩١,٨ ، \overline{ص_2} = ١٩١,٠ ، \overline{ص_3} = ١٩٣,١$$

$$\overline{ص_4} = ٢١٩,٣ ، \overline{ص_5} = ١٨٩,٦ ، \overline{ص_6} = ٢١٣,٦$$

أما الخطأ المعياري ع للفرق بين متوسطين معدلين فيحسب من الصيغة

$$ع' = ع'_{\overline{ص_1} - \overline{ص_2}} = \left[ ١ + \frac{١}{١} \right] \div (١ - ه) = \left[ ١ + \frac{١}{١} \right] \div (١ - ه) = ١,٤١ (١٦)$$

حيث أ' = مجموع المربعات لقيم س محسوبا من بين الأنواع



١، = مجموع المربعات لقيم س محسوبا من خطأ التجريب .

$$ع' = ٩٧,٢١٩ (١ + \frac{٤٥,٨٣}{١١٣,٨٣})$$

بدرجات حرية ١٤

$$= ١٠٥,٠٤٨$$

وبهذه القيمة نكون قادرين على مقارنة ما نريد من أزواج المتوسطات الصادية المعدلة كالمعتاد .

يتمدد تحليل التباين إلى المواقف الأكثر تعقيدا . وكمثال لذلك التجارب ذوات العامل الواحد التي تشتمل على متغيرين ملازمين  $ص_١$  ،  $ص_٢$  يرتبطان خطيا بالمتغير الرئيسي  $ص$  ، حيث يأخذ النموذج الإحصائي الشكل الآتي :

$$ص_{١٢} = \beta_0 + \beta_1 (ص_١ - \bar{ص_١}) + \beta_2 (ص_٢ - \bar{ص_٢}) + \epsilon_{١٢}$$

ويحتاج الأمر هنا إلى تحليل الانحدار المتعدد قبل القيام بتحليل التباين .

## تقنين (١١ - ٢)

بالمجدول الآتي مقادير المحاصيل (ص) لنبات فول الصويا في الفدان ونسبة الإصابة (س) لساق النبات بمرض معين . استخدمت أربعة خطوط أ ، ب ، ح ، و للمقارنة وزرع بكل منها ٤ نباتات . المطلوب (١) توضيح أن هناك علاقة خطية سالبة بين الإصابة بالمرض ومقدار المحصول (٢) توضيح أن معدل المحاصيل يختلف من خط إلى آخر بعد استبعاد أثر الإصابة (٣) إيجاد المتوسطات المعدلة للمحاصيل والمقارنة بين كل زوج منها (٤) توضح أنه إذا لم يُستبعد أثر الإصابة فإن معدل المحاصيل لا يختلف من خط إلى آخر .

الجدول (١١ - أ)

المجموع س ص	الخطوط				القطاعات
	ا س١ ص١	ب س٢ ص٢	ح س٣ ص٣	د س٤ ص٤	
١٠١,٤ ٤٧,٧	٢١,٣ ١٩,٣	٢٨,٣ ١٠,١	٢٦,٧ ٤,٣	٢٥,١ ١٤,٠	(١)
٧٥,٢ ١٤٢,٣	١٩,٧ ٢٩,٢	٢٠,٧ ٣٤,٧	١٤,٧ ٤٨,٢	٢٠,١ ٣٠,٢	(٢)
١٠٨,٦ ٢٨,٥	٢٨,٧ ١,٠	٢٦,٠ ١٤,٠	٢٩,٠ ٦,٣	٢٤,٩ ٧,٢	(٣)
١٢٠,٢ ٢٧,٥	٢٧,٣ ٦,٤	٣٤,١ ٥,٦	٢٩,٠ ٦,٧	٢٩,٨ ٨,٩	(٤)
٤٠٥,٤ ٢٤٦,١	٩٧,٠ ٥٥,٩	١٠٩,١ ٦٤,٤	٩٩,٤ ٦٥,٥	٩٩,٩ ٦٠,٣	المجموع

ح ح ص' = ٦٤٦٥,٩٢٥٦ ، ح ح ص' = ١٠٦٣١,٩٢٢ ، ح ح ص ص' = ٣٥٢,٠٥٨٧

## الفصل الثامن عشر

### الانحدار والارتباط الخطي المتعدد

#### MULTIPLE LINEAR REGRESSION AND CORRELATION

سنعتمد في هذا الفصل على الحاسب الإلكتروني في إجراء الحسابات اللازمة لحل مشكلات الانحدار والارتباط تخففا من الأعباء الحسابية الضخمة التي يتطلبها التحليل . إلا أن هذا لا يعفينا بل يوجب علينا الإحاطة بالأسس والافتراضات والمفاهيم التي يبنى عليها التحليل تحسبا من أخطاء ومزالق التطبيق .

#### أولا : الانحدار الخطي المتعدد

(١٢ - ١) الانحدار الخطي المتعدد كامتداد للانحدار الخطي البسيط :

في تناولنا للانحدار الخطي البسيط في الفصل التاسع من هذا الكتاب افترضنا أن لدينا متغيرا عشوائيا  $y$  نرغب في التنبؤ بقيمه عن طريق قيم متغير غير عشوائي  $x$  على أساس أن العلاقة بين هذين المتغيرين هي علاقة خطية :

$$y = \alpha + \beta x + \epsilon$$

حيث  $\alpha$  ،  $\beta$  بارامتران مجهولان ،

وعلى أساس أنه عند أى قيمة ثابتة  $x$  يكون للمتغير  $y$  توزيع معتدل متوسطه  $\alpha + \beta x$  يتوقف على قيمة  $x$  ، وتباينه عدد ثابت  $\sigma^2$  لا يتوقف على  $x$  .

إلا أنه في كثير من التطبيقات يرى الباحث أن استخدام متغير واحد  $\mu$  لا يصلح للتنبؤ بقيمة المتغير  $\mu$  بدقة كافية ، ويأمل أن يحصل على تنبؤات أكثر دقة إذا استخدم عدة متغيرات  $\mu_1, \mu_2, \dots, \mu_p$  ، يشعر بخبرته في ميدان عمله أن لها دورا في عملية التنبؤ . وإذا تبيننا افتراضات مماثلة لتلك التي تبينناها في الانحدار الخطي البسيط من أن هذه المتغيرات غير عشوائية وترتبط بالمتغير العشوائي  $\mu$  بعلاقة خطية :

$$\mu = \alpha + \beta_1 \mu_1 + \beta_2 \mu_2 + \dots + \beta_p \mu_p \quad (1)$$

حيث  $\alpha, \beta_1, \beta_2, \dots, \beta_p$  بارامترات مجهولة ، ومن أنه عند أي مجموعة ثابتة  $\mu_1, \mu_2, \dots, \mu_p$  ، سيكون للمتغير  $\mu$  توزيع معتدل ذو تباين ثابت  $\sigma^2$  ، فإن الانحدار يسمى في هذه الحال بالانحدار الخطي المتعدد أو اختصارا بالانحدار الخطي كما تسمى البارامترات  $\beta_1, \beta_2, \dots, \beta_p$  بمعاملات الانحدار الجزئية **partial regression coefficients** ، وتكون دراسة هذا الانحدار مجرد تعميم للمفاهيم والأساليب والصيغ والاختبارات المستخدمة في الانحدار الخطي البسيط ولا تختلف عنها إلا في درجة المشقة في تناول البيانات ، كما تختلف بطبيعة الحال في درجات الحرية التي تعتمد على عدد الثوابت في معادلة الانحدار التي تقدر من العينة . وإذا كان القارئ قد اطمأن لمعرفته بأسس وأساليب الانحدار الخطي البسيط فسوف لا يجد صعوبة تذكر في تناول الانحدار الخطي المتعدد كما يتبين من الفقرات الآتية .

#### (أ) إيجاد معادلة الانحدار :

في الانحدار الخطي البسيط نقوم بتجربة نحصل منها على عينة من  $n$  من المشاهدات ( $\mu, \mu_1$  ،  $\mu_2$  ،  $\dots$  ،  $\mu_p$ ) تأخذ الشكل الآتي :

$$\begin{array}{cccc} s_1 & s_2 & \dots & s_n \\ v_1 & v_2 & \dots & v_n \end{array}$$

وتستخدم هذه البيانات في تقدير البارامترين المجهولين  $\alpha$  ،  $\beta$  بقيمتين  $1$  ،  $b$  توظفة لكتابة معادلة الانحدار :

$$v = 1 + b s$$

التي تمثل « أحسن خط » يلائم تلك المشاهدات من وجهة نظر مبدأ المربعات الصغرى ، الذى يستلزم كما نعلم تحديد القيمتين  $\alpha$  ،  $\beta$  اللتين تجعلان الدالة :

$$e = (v - \hat{v})^2 = (v - (\alpha + \beta s))^2$$

نهاية صغرى . وتحقيق هذا الغرض يستلزم بدوره إجراء عملية التفاضل الجزئى بالنسبة إلى كل من  $\alpha$  ،  $\beta$  ومساواة كل من الناتجين بالصفر للحصول على معادلتين خطيتين فى هذين المجهولين تسميان بالمعادلتين المتعادتين وهما :

$$e = \alpha + \beta s = v$$

$$e = \alpha + \beta s = v$$

وحل هاتين المعادلتين معا يعطينا القيمتين  $1$  ،  $b$  المطلوبتين .

وهذه الخطوات هى بذاتها الخطوات التى تتخذ عند تناول الانحدار الخطى المتعدد . إذ نقوم بتجربة نحصل منها على عينة من  $n$  من المشاهدات  $(s_1, v_1, s_2, v_2, \dots, s_n, v_n)$  ، حيث  $s = 1, 2, \dots, n$  ،  $v$  تأخذ الشكل المبين بالجدول (١٢ - ١) الآتى :

المجلد (١٢ - ١)

بيانات الانحدار الخطي المتعدد

الملاحظات	ص	س <sub>١</sub>	س <sub>٢</sub>	....	س <sub>ك</sub>
(١)	ص <sub>١</sub>	س <sub>١١</sub>	س <sub>٢١</sub>	...	س <sub>ك١</sub>
(٢)	ص <sub>٢</sub>	س <sub>١٢</sub>	س <sub>٢٢</sub>	...	س <sub>ك٢</sub>
...	...	...	...	...	...
(٣)	ص <sub>٣</sub>	س <sub>١٣</sub>	س <sub>٢٣</sub>	...	س <sub>ك٣</sub>
...	...	...	...	...	...
(٥)	ص <sub>٥</sub>	س <sub>١٥</sub>	س <sub>٢٥</sub>	...	س <sub>ك٥</sub>

ولتقدير البارامترات المجهولة  $\alpha$  ،  $\beta_1$  ،  $\beta_2$  ، ... ،  $\beta_K$  من هذه البيانات نستخدم مبدأ المربعات الصغرى لإيجاد القيم  $a$  ،  $b_1$  ،  $b_2$  ، ... ،  $b_K$  بواسطة كتابة معادلة الانحدار :

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + \dots + b_K x_K \quad (2)$$

وتحقيق هذا المبدأ يستلزم أن تؤخذ القيم  $a$  ،  $b_1$  ، ... ،  $b_K$  بحيث تكون الدالة

$$S = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_n - \hat{y}_n)^2$$

نهاية صغرى ، وهذا يستلزم بدوره إجراء عملية التفاضل الجزئي بالنسبة إلى كل من  $\alpha$  ،  $\beta_1$  ،  $\beta_2$  ، ... ،  $\beta_K$  ومساواة النتائج بالصفر للحصول على  $K + 1$  من المعادلات الخطية في هذه المجاهيل تسمى بالمعادلات المعتادة . وحل هذه المعادلات معا يعطينا التقديرات المطلوبة  $a$  ،  $b_1$  ،  $b_2$  ، ... ،  $b_K$  وعددها  $K + 1$  .

ولنا أن نتصور هنا مدى المشقة التي نلقاها في حساب المجاميع ومجاميع المربعات ومجاميع حواصل الضرب المطلوبة لوضعها في هذه المعادلات ثم حل المعادلات ذاتها ، خاصة إذا كان عدد المتغيرات السينية أكبر من اثنين . وحتى في حالة وجود متغيرين  $x_1, x_2$  ، حيث نستهدف الوصول إلى معادلة الانحدار

$$\hat{y} = a + b_1 x_1 + b_2 x_2$$

يكون علينا حل المعادلات المعتادة الآتية :

$$\sum y = n a + b_1 \sum x_1 + b_2 \sum x_2$$

$$\sum x_1 y = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \quad (1)$$

$$\sum x_2 y = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2$$

وهذا الحل يحتاج أولاً إلى حساب ثمانية مجاميع هي  $\sum y, \sum x_1, \sum x_2, \sum x_1^2, \sum x_2^2, \sum x_1 x_2, \sum x_1 y, \sum x_2 y$  ، كثير من الجهد والوقت . ولذلك نلجأ إلى الحاسب الإلكتروني ليقوم عنا بكل هذه العمليات ويعطينا التقديرات المطلوبة بكل سرعة ودقة .

(ب) إيجاد الخطأ المعياري لتقدير  $\hat{y}$  من خط الانحدار :

في الانحدار الخطي البسيط استخدمنا مقياساً رأينا أهميته البالغة في عمليات الاستنتاج الإحصائي هو الخطأ المعياري للتقدير الذي رمزنا له بالرمز  $s_{\hat{y}}$  وعرفناه في البند (٩ - ٤) بأنه الانحراف المعياري للقيم المعيارية المشاهدة  $y$  حول القيم  $\hat{y}$  المقدرة من معادلة الانحدار أى عرفناه كالآتي :

$$s_{\hat{y}} = \frac{1}{n-2} (\text{مجموع مربعات الأخطاء})$$

وحيث حللنا الاختلاف الكلي في القيم الصادية وهو  $\epsilon$  (صـ - ص') إلى مركبتين  
 تعبر الأولى وهي  $\epsilon$  (صـ - ص') عن الاختلاف الناشئ عن الانحدار ( الاختلاف  
 المفسر ) وتعبر الثانية وهي  $\epsilon$  (صـ - ص') عن الاختلاف المتبقى الناشئ عن  
 الانحراف عن خط الانحدار ( الاختلاف غير المفسر ) أمكننا كتابة مربع الخطأ  
 المعياري للتقدير كالآتي :

$$\epsilon^2_{\text{م.م.}} = \epsilon^2_{\text{ص.ص.}} - \epsilon^2_{\text{ص.ص.}} = \frac{1}{n-2} \times \text{الاختلاف غير المفسر}$$

بنفس المنطق نستخدم نفس التعريف في حالة الانحدار الخطي المتعدد مع مراعاة  
 درجات الحرية ، فنرمز للخطأ المعياري لتقدير ص من خط الانحدار بالرمز  
 $\epsilon_{\text{ص.ص.}} \dots \epsilon_{\text{ص.ص.}} \dots \epsilon_{\text{ص.ص.}}$  حيث :

$$(4) \quad \epsilon^2_{\text{ص.ص.}} \dots \epsilon^2_{\text{ص.ص.}} \dots \epsilon^2_{\text{ص.ص.}} = \frac{1}{(n-k)-1} \times \text{الاختلاف غير المفسر}$$

حيث  $n$  حجم العينة ،  $k$  عدد المتغيرات السينية ومع ملاحظة أن  $k + 1$  هو  
 عدد الثوابت في معادلة الانحدار التي تقدر من العينة .

ملاحظة (١) :

ليس من المناسب هنا تقديم الصيغة العامة للخطأ المعياري للتقدير وبكفينا أن  
 نفهم المعنى الذي يتضمنه ، أما حساب قيمته فتتركه للحاسب الإلكتروني . إلا  
 أنه حين يكون هناك متغيران تبؤيان اثنان فقط ( $k = 2$ ) فيمكن إثبات أن

$$\epsilon^2_{\text{ص.م.}} = \frac{1}{n-3} (\epsilon_{\text{ص.ص.}}^2 - \epsilon_{\text{ص.ص.}}^2 - \epsilon_{\text{ص.ص.}}^2 - \epsilon_{\text{ص.ص.}}^2 - \epsilon_{\text{ص.ص.}}^2) \quad (5)$$



وهذه الصيغة هي امتداد للصيغة (١٠) بالبند (٩ - ٤) التي تتناول الحالة التي يكون لدينا فيها متغير تنبؤى واحد .

هذا مع ملاحظة أن الصيغة العامة للخطأ المعياري تكتب بأسلوب رياضي يتطلب معرفته دراسة مسبقة لموضوع المصفوفات ، والواقع أن الدراسة النظامية للانحدار المتعدد تعتمد برمتها على هذا الأسلوب غير أننا في التطبيق العملي لا نحتاج إليه .

### (ج) اختبار دلالة الانحدار ككل :

في الانحدار الخطى البسيط اهتممنا باختبار دلالة الانحدار أى باختبار وجود علاقة خطية بين المتغير التابع  $y$  والمتغير المستقل  $x$  وهذا يكافئ اختبار الفرض الصفرى  $\beta = 0$  ضد الفرض  $\beta \neq 0$  . واستخدمنا لهذا الغرض اختبار  $t$  بالصيغة (١١) بالبند (٩ - ٥ - أولاً) مع وضع  $\beta = 0$  . أو اختبار  $F$  بالصيغة (٢١) بالبند (٩ - ٧) أو بالصيغة المكافئة (٢٤) ، بشرط توفر شروط الانحدار .

كذلك يهتما في حالة الانحدار الخطى المتعدد اختبار وجود علاقة خطية بين المتغير التابع  $y$  والمتغيرات المستقلة  $x_1, x_2, \dots, x_p$  ، وهذا الاختبار يكافئ اختبار الفرض الصفرى  $\beta_1 = \beta_2 = \dots = \beta_p = 0$  . وفي هذه الحالة نستخدم أى من الصيغتين (٢١) أو (٢٤) المذكورتين مع مراعاة استخدام درجات الحرية المناسبة . وهاتان الصيغتان هما :

$$F = \frac{\text{التباين المقسّر}}{\text{التباين غير المقسّر}} \quad \text{بدرجتي حرية ك ، ن - ك - ١} \quad (٦)$$

$$\text{أو } F = \frac{r^2 / ك}{(١ - r^2) / (ن - ك - ١)} \quad \text{بدرجتي حرية ك ، ن - ك - ١} \quad (٧)$$

حيث  $n$  حجم العينة ،  $k$  عدد المتغيرات السينية ،  $r^2$  هو معامل التحديد الذى يعبر كالمعاد عن نسبة الاختلاف فى المتغير  $y$  التى يفسرها خط الانحدار وسنرمز له هنا بالرمز  $r^2$  (٢١...٥) حيث

$$(٨) \quad r^2 = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}} \quad (٢١...٥)$$

(٥) اختبار دلالة معاملات الانحدار الجزئية :

فى الانحدار الخطى البسيط يكون لدينا معامل انحدار واحد  $\beta$  ، ويمكن أن نختبر ما إذا كان هذا المعامل يأخذ قيمة ما نفترضها بواسطة الإحصاءة (١١) المقدمة بالبند (٩ - ٥ - أولاً) وهى

$$(٩) \quad \frac{\beta - B}{\sigma_{\beta}} = \frac{\beta - B}{\sigma_{\beta} / \sqrt{22}} = t$$

التي يكون لها توزيع  $t$  بدرجات حرية  $n - ٢$  بشرط تحقق افتراضات الانحدار . ويلاحظ أن مقام هذا الكسر وهو  $\sigma_{\beta}$  (ب) =  $\sigma_{\beta} / \sqrt{22}$  هو تقدير للانحراف المعياري للمتغير العشوائى  $B$  الذى تعتبر  $B$  إحدى قيمه المشاهدة - وحيث  $B$  تقدير معامل الانحدار  $\beta$  الذى نجده من العينة . والمتغير  $B$  هو متغير معتدل وسطه الحسابى  $\beta$  وانحرافه المعياري  $\sigma$  (ب) يقدر من العينة بالمقدار  $\sigma_{\beta}$  (ب) .

أما فى الانحدار الخطى المتعدد فلدينا  $k$  من معاملات الانحدار الجزئية  $\beta_1, \beta_2, \dots, \beta_k$  (بالإضافة إلى البارامتر  $\alpha$ ) وبهنا أن نختبر ما إذا كانت هذه المعاملات تأخذ قيما معينة . ونفرض منطق البند (٩ - ٥) نعتبر أن القيمة  $\beta_i$

(١ = ٢ ، ... ، ك) التى نحصل عليها من عينة لتقدير البارامتر  $\beta_j$  ، هى إحدى القيم المشاهدة من متغير عشوائى  $B_j$  ذى توزيع معتدل متوسطه  $\beta_j$  وانحرافه المعيارى  $\sigma$  (ب<sub>ج</sub>) نقدره بقيمة سنرمز لها بالرمز  $\hat{\sigma}_j$  وبالتالي يكون للإحصاءة

$$T_j = \frac{\beta_j - B_j}{\hat{\sigma}_j} \quad \text{ع (ب<sub>ج</sub>)} \quad \text{ك} = ١ ، ٢ ، ... ، ك \quad (١٠)$$

توزيع ت بدرجات حرية ك - (١ + ك) بشرط تحقق افتراضات الانحدار . وهذه الإحصاءة تصلح لاختبار أن يأخذ أى معامل  $\beta_j$  أى قيمة نفترضها ، وبصفة خاصة لاختبار الفرض  $\beta_j = ٠$  لأهمية ذلك فى بحث مدى مساهمة المتغير المناظر  $X_j$  فى التنبؤ من معادلة الانحدار .

#### ملاحظة (٢) :

ليس من المناسب هنا أيضا تقديم الصيغة العامة التى نوجد منها التقدير  $\hat{\sigma}_j$  (ب<sub>ج</sub>) للانحراف المعيارى للمتغير  $B_j$  وسنعمد فى ذلك أيضا على الحاسب الالكترونى ويكفي أن نخطط بالدور الذى يقوم به . إلا أنه حين يكون هناك متغيران تنبؤيان اثنان فقط (ك = ٢) فيمكن إثبات أن هذا التقدير يحسب من الصيغة الآتية :

$$\hat{\sigma}_j \text{ (ب<sub>ج</sub>)} = \frac{\sum_{i=1}^n \frac{X_{ij}^2}{\sqrt{(1 - r_{jj}^2) \cdot (n - 2)}}}{\sqrt{(n - 2) \cdot (1 - r_{jj}^2)}} \quad \text{ك} = ١ ، ٢ \quad (١١)$$

حيث  $\sum_{i=1}^n \frac{X_{ij}^2}{\sqrt{(1 - r_{jj}^2) \cdot (n - 2)}}$  =  $\frac{1}{n - 2} \times$  الاختلاف غير المفسر

هو مربع الخطأ المعيارى للتقدير ويمكن حسابه من الصيغة (٥) ، وحيث  $r_{jj}^2$  هو معامل التحديد للمتغيرين  $X_j$  ،  $X_p$  ويمكن حسابه من الصيغة :

$$(١٢) \quad \frac{[ (٢,٣,٤) - (١,٢,٣) ]}{(٢,٣) - (١,٢)} = ١,٣$$

كما أن  $(٢,٣)$  هو مجموع المربعات للمتغير  $٢$  ،  $(١,٢)$  هو مجموع المربعات للمتغير  $١$  .

## (١٢ - ٢) استخدام الحاسب الالىكترونى :

يتضح من البند السابق أن حل المشكلات العملية فى تحليل الانحدار يتطلب القيام بعمليات حسابية طويلة سواء فى إيجاد المجاميع أو حل المعادلات أو فى استخراج تقديرات للأخطاء المعيارية أو فى استخراج قيم  $t$  أو قيم  $F$  اللازمة للاستنتاج الإحصائى ، مما يستغرق جهدا شاقا ووقتا طويلا ، إضافة إلى صعوبة تلافى أخطاء هذه الحسابات إذا أجريت يدويا . ولذلك فإن معظم المشكلات التطبيقية تستعين بالحاسبات الالىكترونية التى تنوب عنا فى استخراج ما نريد بكل سرعة ودقة وتترك لنا فقط مهمة تفسير النتائج واتخاذ القرارات بناء على ما قدمته من معلومات .

واستخدام هذه الحاسبات لا يستلزم أن يكون الباحث قادرا على تشغيلها بنفسه بل يكفيه أن يتصل بأحد مراكز الحاسبات وتقديم مصغوفة البيانات التى حصل عليها من التجربة مع تحديد مجموعة الأعداد التى تخص كلا من المتغير الصادى والمتغيرات السنية ، وتحديد ما يريد إنجاده من معلومات .

وهناك كثير من الحاسبات تحتزن برامج إحصائية تقوم بكافة أنواع التحليل ، ومن هذه البرامج ما يلى :

### STATPACK, MINITAB, COSAP AND SPSS

ومثل هذه البرامج من شأنها توفير الجهد والوقت مع ضمان الدقة والأمان ولذلك يعتمد هذا الفصل فى حساباته وتحليلاته على الحاسبات الالىكترونية التى أصبحت

اليوم في متناول الجميع . وسنوضح ذلك بالمثال الآتي الذى يتناول حالة متغيرين  
 تنبؤين  $\hat{y}_1$  ،  $\hat{y}_2$  . على أن الأسلوب المستخدم يمتد إلى الحالات التى تتناول  
 أكثر من متغيرين دون أى تعديل .

مثال (١٢ - ١) :

أرادت إحدى الشركات التجارية الكبيرة أن تعد طريقة للتنبؤ بعدد الوحدات  
 التى تباع في الشهر في أى فرع من فروعها العشرة ، على افتراض أن الوحدات  
 المباعة في أى فرع تتأثر بعاملين هما (١) عدد البائعين في الفرع ، (٢) مقدار  
 ما يصرفه الفرع شهريا على الإعلان . وتحقيقا لهذا الغرض جمعت بيانات من  
 الفروع ودونت في الجدول (١٢ - ٣) الآتي .

الجدول (١٢ - ٣)

الفرع	عدد الوحدات المباعة ص	عدد العاملين $x_1$	تكاليف الاعلان $x_2$
(١)	٢٥	٥	١٠
(٢)	٢٠	٢	١١
(٣)	٣٠	٦	١٢
(٤)	٢٥	٤	١٣
(٥)	٢٥	٣	١٤
(٦)	٢٢	٦	١٥
(٧)	٢٥	٤	١٢
(٨)	٢١	٣	١١
(٩)	٢٠	٢	١٠
(١٠)	٢٧	٥	١٢

المطلوب إيجاد معادلة الانحدار الخطي للمتغير ص على المتغيرين س<sub>١</sub> ، س<sub>٢</sub> واختبار دلالة الانحدار ككل ودلالة كل من معاملي الانحدار .

الحل :

للتوضيح سنقوم بحل هذه المسألة مرتين : باستخدام الحاسب ثم يدونه .

( أولا ) الحل باستخدام الحاسب :

أدخلت البيانات المدونة بالجدول (١٢ - ٣) في حاسب الكروني فأخرج المعلومات المدونة بالجدول (١٢ - ٤) الآتي ، مع ملاحظة أن عدد المتغيرات التنبؤية ك = ٢ وأن حجم العينة ن = ١٠ .

الجدول (١٢ - ٤)

مخرجات الحاسب الالكتروني لبيانات المثال (١٢ - ١)

(1)	Regress.of	Y	Number of Units Sold		
(2)	on	X <sub>1</sub>	Number of Salespeople		
(3)		X <sub>2</sub>	Amount of Advertising Expenditure		
(4)	Variable Name	Regress. Coeff	S.E. of Coeff.	$\bar{X}$	D.F.
(5)	Constant	6.85263			
(6)	X <sub>1</sub>	2.10526	.17708	11.88891	7
(7)	X <sub>2</sub>	.87719	.16125	5.42652	7
(8) Coefficient of Determination ( $R^2$ ) = .974659					
(9) Estimated Standard Error of Estimate = .722013					
Analysis of Variance for Regression					
(10)	Source of Variation	SS	D.F.	MS	F
(11)	Regression	140.351	2	70.1754	134.615
(12)	Residual	3.64912	7	.521303	

### (أ) معادلة الانحدار :

بالتأمل في الجدول (١٢ - ٤) نرى أن الحاسب قد قام بإيجاد التقديرات المطلوبة للبارامترات  $\alpha$  ،  $\beta$  ،  $\beta$  ، وهي :

$$1 = 6,05263 ، ب = 2,10526 ، ب = 0,87719$$

وهذه القيم الثلاث هي تلك المدونة بالصفوف ٥ ، ٦ ، ٧ من العمود الثاني بالجدول . وإذن معادلة الانحدار هي - التقريب إلى ٣ خانات عشرية :

$$\hat{س} = 6,053 + 2,105 س + 0,877 س$$

فمثلا إذا كان عدد البائعين في الفرع  $س = 4$  والتكاليف الشهرية  $س = 12$  فإننا نتوقع أن يكون عدد الوحدات المباعة حوالي ٢٤ وحدة .

### (ب) اختبار دلالة الانحدار الخطي :

من الجدول (١٢ - ٤) نجد أن الحاسب قد حسب لنا جدول التباين في السطور الثلاثة الأخيرة ، فقد أعطانا كلا من :

$$\text{الاختلاف المفسر (الانحدار)} = 140,351 \text{ بدرجات حرية ك} = 2$$

$$\text{الاختلاف غير المفسر ( الانحراف عن خط الانحدار )} = 3,64912 \text{ بدرجات}$$

$$\text{حرية ه} - \text{ك} = 1 - 2 = 7 \text{ بل أعطانا أيضا نسبة التباين في مستخدما الصيغة (٦)}$$

وهي

$$ف = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} = \frac{140,1704}{0,021303} = 134,615$$

### الاستنتاج :

لما كانت القيمة ١٣٤,٦١٥ تزيد كثيرا عن القيمة الحرجة ف  $\gamma$  ،  $\gamma_{1,0.1}$  ٩,٥٥=

نرفض الفرض الصفري أن  $\beta_1 = \beta_2 = 0$  عند مستوى عالى من الدلالة بمعنى أن هناك علاقة خطية بين عدد الوحدات المباعة من ناحية وعدد البائعين وتكاليف الإعلان من ناحية أخرى .

كما أن الحاسب قد أعطانا قيمة معامل التحديد بين المتغير ص والمتغيرين  $S_1$  ،  $S_2$  وهو  $S^2 = 0.974609$  (ص ١٠٢) وهذا يعنى أن المتغيرين التنبؤيين قد فسرا حوالى ٩٧,٥٪ من التغير في ص وهذا جزء كبير يدعم الحكم بوجود العلاقة الخطية .

### (ج) اختبار دلالة كل من معامل الانحدار :

أعطانا الحاسب في السطرين السادس والسابع من العمود الثالث التقديرين الخاصين بالخطأ المعياري لمعامل الانحدار الجزئيين وهما  $E (b_1) = 0.17708$  ،  $E (b_2) = 0.16125$  ، كما أعطانا في العمودين الآخرين من هذين السطرين قيمتي  $t_1$  ،  $t_2$  مستخدما الصيغة (١٠) بعد وضع  $\beta_1 = 0$  ،  $\beta_2 = 0$  وهما  $t_1 = 11.88891$  ،  $t_2 = 5.42652$  ولكل منهما سبع درجات حرية .

### الاستنتاج :

نظرا لأن  $t_{0.01, 17} = 3.499$  نرفض كلا من الفرضين الصفريين  $\beta_1 = 0$  ،  $\beta_2 = 0$  عند مستوى الدلالة ٠,٠١ وهذا يعنى أن كلا من المتغيرين التنبؤيين يسهم إسهاما جوهريا في التنبؤ بعدد الوحدات المباعة .

نلخص ما وجدناه في هذه التجربة كما يلي : بناء على البيانات المشاهدة في العينة على أساس أن افتراضات الانحدار متحققة ، تأخذ معادلة الانحدار الشكل الآتى :

$$\hat{V} = 6.053 + 2.105 S_1 + 0.877 S_2$$



وهذه المعادلة تعبر تعبيراً مناسباً عن العلاقة الحقيقية بين المتغير  $\alpha$  ( عدد الوحدات المباعة ) والمتغيرين  $\alpha_1$  ،  $\alpha_2$  ( عدد البائعين وقيمة تكاليف الإعلان ) ، وتفسر حوالى ٩٧,٥٪ من التغير في قيم  $\alpha$  . كما أن كلا من هذين المتغيرين يسهم إسهاماً جاداً في التنبؤ بقيم هذا المتغير .

( ثانياً ) الحل بغير استخدام الحاسب :

إذا لم يكن الحاسب الآلى متوفراً وكان الانحدار ذا متغيرين تنبؤيين فقط ، يمكن إجراء العمليات الحسابية المطلوبة باستخدام حاسب الجيب دون تحمل مشقة كبيرة . وفي هذا المثال يتخذ الحل الخطوات الآتية :

(١) إيجاد معادلة الانحدار :

نبدأ بحساب المجاميع الآتية

$\sum \alpha = 250$  ،  $\sum \alpha_1 = 6394$  ،  $\sum \alpha_2 = 40$  ،  $\sum \alpha_1^2 = 180$  ،  
 $\sum \alpha_2^2 = 120$  ،  $\sum \alpha_1 \alpha_2 = 1464$  ،  $\sum \alpha_1 \alpha = 489$  ،  $\sum \alpha_2 \alpha = 1050$  .  
 ثم نعوض بهذه القيم في المعادلات المعتادة (٣) كالآتي :

$$10 \alpha + \beta_1 120 + \beta_2 40 = 250$$

$$40 \alpha + \beta_1 489 + \beta_2 180 = 1050$$

$$120 \alpha + \beta_1 1464 + \beta_2 489 = 3040$$

وهناك عدة طرق لحل مثل هذه المعادلات نختار منها هنا طريقة دوليتل Doolittle التي تحول مصفوفة المعاملات والتوابت في المعادلات المعتادة إلى مصفوفة مثلثية تكون جميع عناصر القطر الرئيسى فيها مساوية للواحد . ولهذه الطريقة روتين خاص من التعليمات تبين من الجدول (١٢ - ٥) الآتي .

المجدول (١٢ - ٥)

طريقة دوليتل لحل المعادلات الخطية

التعليمات	الصف	$\alpha$	$\beta_1$	$\beta_2$	الثوابت
معاملات وثابت للمعادلة الأولى	$1^2$	١٠	٤٠	١٢٠	٢٥٠
معاملات وثابت للمعادلة الثانية	$2^2$	٤٠	١٨٠	٤٨٩	١٠٥٠
معاملات وثابت للمعادلة الثالثة	$3^2$	١٢٠	٤٨٩	١٤٦٤	٣٠٤٠
$1^2$	$1^2$	١٠	<u>٤٠</u>	<u>١٢٠</u>	٢٥٠
$1^2 \div 1^2$	$1^2$	١	٤	١٢	٢٥
$2^2 - 1^2 \times 4$	$2^2$	٠	٢٠	<u>٩</u>	٥٠
$2^2 \div 2^2$	$2^2$	٠	١	٠,٤٥	٢,٥
$3^2 - 1^2 \times 10 - 2^2 \times 4$	$3^2$	٠	٠	١٩,٩٥	١٧,٥
$3^2 \div 3^2$	$3^2$	٠	٠	١	٠,٨٧٧١٩

يلاحظ في هذه التعليمات أن هناك ثلاثة مقادير ثابتة هي  $1^2 = ٤٠$  ،  $2^2 = ١٢٠$  ،  $3^2 = ٩$  وهذه تسمى محاور الارتكاز . أما بقية القيم التي تشتمل على الرمز  $s$  فتتغير حسب قيمة  $s$  التي تعبر عن رقم العمود ( $s = ١, ٢, ٣$  ، ٤) فمثلا  $1^2 s$  تعبر عن العنصر الذي بالصف  $1^2$  والعمود  $s$  وبهذا يكون

$$1^2 = ٤٠ ، 1^2 s = ١٨٠ ، 2^2 = ٤٨٩ ، 2^2 s = ١٠٥٠ ، 3^2 = ١٤٦٤ ، 3^2 s = ٣٠٤٠$$

تبدأ طريقة دولتيل بتدوين معاملات وثوابت المعادلات كما هو مبين بالصفوف الثلاثة الأولى المشار إليها بالرموز  $\alpha, \beta, \gamma$  . ثم تحسب عناصر الصفوف التالية الواحد بعد الآخر باستخدام التعليمات المبينة أمامه .

فالصف الرابع  $\delta$  يتكون من نفس عناصر الصف  $\gamma$  . والصف التالي  $\epsilon$  يتكون بقسمة عناصر الصف السابق  $\gamma$  على معامل  $\alpha$  وهو ١٠ ، والصف التالي  $\zeta$  تتكون عناصره باستخدام التعليمات المبينة أمامه كالاتي :

وبوضع  $\gamma = 1$  نجد أن العنصر الأول في هذا الصف هو  $40 - 1 \times 40 = 0$  .

وبوضع  $\gamma = 2$  نجد أن العنصر الثاني في هذا الصف هو  $180 - 2 \times 40 = 100$  .

وبوضع  $\gamma = 3$  نجد أن العنصر الثالث في هذا الصف هو

$$9 = 12 \times 40 - 489$$

وبوضع  $\gamma = 4$  نجد أن العنصر الرابع في هذا الصف هو

$$50 = 25 \times 40 - 1050$$

ويلاحظ أننا استخدمنا هنا محور الارتكاز  $\gamma = 40$  .

بنفس الطريقة نوجد بقية الصفوف الثلاثة . وبذلك تتحول المصفوفة التي رمز لصفوفها بالرموز  $\alpha, \beta, \gamma, \delta, \epsilon, \zeta$  وهي المصفوفة الأصلية إلى المصفوفة المثلثية التي رمز لصفوفها بالرموز  $\alpha, \beta, \gamma, \delta, \epsilon, \zeta$  وبالتالي تتحول مجموعة المعادلات المعتادة إلى المجموعة المكافئة الآتية :

$$\alpha + \beta_1 12 + \beta_2 20 = \gamma_1 \quad (\text{من الصف } \alpha)$$

$$\beta_1 + \beta_2 0.45 + \beta_3 2.5 = \gamma_2 \quad (\text{من الصف } \beta)$$

$$\beta_2 + \beta_3 0.87719 = \gamma_3 \quad (\text{من الصف } \gamma)$$

من هذه المعادلات ينتج أن القيم  $\alpha, \beta, \gamma$  التي تحقق هذه المعادلات معا هي :

$$٠,٨٧٧١٩ = ب$$

$$٠,٨٧٧١٩ \times ٠,٤٥ - ٢,٥ = ب \quad ٠,٤٥ - ٢,٥ = ب \therefore$$

$$٢,١٠٥٢٦ =$$

$$١, \quad ٢٥ - ١٢ = ب \quad ٤ - ب =$$

$$٢,١٠٥٢٦ \times ٤ - ٠,٨٧٧١٩ \times ١٢ - ٢٥ =$$

$$٦,٠٥٢٦٨ =$$

فتكون معادلة الانحدار مقربة إلى ثلاث خانات عشرية هي :

$$ص = ٠,٨٧٧ + ب \quad ٢,١٠٥ + ب \quad ٦,٠٥٣ = ص$$

وهي نفس المعادلة التي أوجدناها لنا الحاسب الآلي .

(ب) اختبار دلالة الانحدار :

نحتاج هنا إلى تحليل الاختلاف في ص إلى مركبتين كالآتي :

$$\frac{٢(ص)}{٥} - ص = ص \quad ٢(ص) = ص \quad ص - \frac{٢(ص)}{٥}$$

$$٩ = ١ - ٥ = \frac{٢(٢٥٠)}{١٠} - ٦٣٩٤ = ١٤٤ \text{ بدرجات حرية } ٩$$

من الصيغة (٥) نجد أن :

الاختلاف غير المفسر = ص - أ - ب - ج - د - هـ - ز - ح - ط - ص

$$٣٠٤٠ \times ٠,٨٧٧١٩ - ١٠٥٠ \times ٢,١٠٥٢٦ - ٢٥٠ \times ٦,٠٥٢٦٨ - ٦٣٩٤ =$$

$$٧ = ١ - ٥ - ٦ - ٧ - ٨ - ٩ - ١٠ = ٣,٦٤٩٤ =$$

$$\therefore \text{الاختلاف المفسر} = 144 - 3,6494$$

$$140,3506 = \text{بلوجات حرية ك} = 2$$

لاختبار دلالة الانحدار نستخدم الصيغة (٦) كالآتي :

$$F = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}} = \frac{2 \div 140,3506}{7 \div 3,6494} = 134,6049 =$$

وهذه هي نفس القيمة التي أوجدناها الحاسب (مع فارق التقريب) وهي كما ذكرنا ذات دلالة عالية وتدعونا لرفض الفرض الصفري  $\beta_1 = \beta_2 = 0$ .

$$\text{معامل التحديد } r^2 \text{ من (١, ٢)} = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلي}} = \frac{140,3506}{144} = 0,97466 =$$

وهي نفس القيمة التي أوجدناها الحاسب .

(ج) اختبار دلالة كل من معامل الانحدار :

نحتاج هنا أولاً إلى تقدير الخطأ المعياري لكل من المعاملين بالصيغة (١١) وهي

$$e(B_i) = \frac{\sqrt{\frac{e}{n-2}}}{\sqrt{(1 - r^2) \cdot (22 - 1)}} \quad e = 1, 2$$

ثم استخدام اختبارات لكل منهما بالصيغة (١٠) وهي

$$t_i = \frac{B_i - 0}{e(B_i)} \quad e = 1, 2$$

بدرجات حرية عددها ن - ك - ١ = ٧ .

$$\text{لدينا : } \frac{1}{1 - 0.9 - 0.1} \times \text{الاختلاف غير المفسر} = 0.5212427$$

$$0.5212427 = 3.6494 \times \frac{1}{\sqrt{V}} =$$

$$0.7220 = \text{ع. ص. س. س.}$$

$$20 = \frac{140}{10} - 180 = (1, 2, 3)$$

$$24 = \frac{120}{10} - 1464 = (1, 2, 3)$$

$$9 = \frac{120 \times 40}{10} - 489 = (1, 2, 3)$$

$$0.16875 = \frac{81}{24 \times 20} = 1.1$$

$$0.83125 = 1 - 1.1$$

$$0.17707 = \frac{0.72198}{\sqrt{20 \times 0.83125}} = \text{منها ع (ب)}$$

$$0.16164 = \frac{0.72198}{\sqrt{24 \times 0.83125}} = \text{ع (ب)}$$

$$\text{بدرجات حرية } 7 \quad 11.888 = \frac{2.100}{0.17707} = \text{ت. 1}$$

$$\text{بدرجات حرية } 7 \quad 0.4206 = \frac{0.877}{0.16164} = \text{ت. 1}$$

وهاتان القيمتان أعطاهما الحاسب وقد رأينا أن كليهما ذو دلالة عالية وتدعوان إلى رفض كل من الفرضين  $\beta_1 = 0$  ،  $\beta_2 = 0$  عند مستوى الدلالة ٠,٠١ .

## (١٢ - ٣) أسلوب آخر لاختبار دلالة معاملات الانحدار الجزئية :

لنبداً بالحالة التي يكون لدينا فيها متغيران  $S_1$  ،  $S_2$  . إذا أوجدنا الاختلاف المفسر الناشئ عن انحدار هذين المتغيرين على المتغير  $S_3$  حين يستخدمان معا ( بدرجتين من درجات الحرية ) ثم أوجدنا الاختلاف الناشئ عن انحدار المتغير  $S_1$  على المتغير  $S_3$  حين يستخدم وحده ( بدرجة واحدة من الحرية ) فإن الفرق بين هذين الاختلافين هو مقدار الاختلاف المفسر الذي ساهم به المتغير  $S_1$  عند ضمه إلى المتغير  $S_2$  . أي هو المساهمة الخاصة بالمتغير  $S_1$  بعد استبعاد مساهمة  $S_2$  ويمكن اختبار الفرض الصفري  $\beta_1 = 0$  ضد الفرض  $\beta_1 \neq 0$  بواسطة اختبار ف بالصيغة (٦) كالآتي :

$$F = \frac{\text{التباين المفسر (للمتغير } S_1 \text{ بعد استبعاد } S_2 \text{)}}{\text{التباين غير المفسر}}$$

$$(13) \quad = \frac{[ \text{انحدار } S_1 , S_2 - ( \text{انحدار } S_2 ) ]^2 / (1)}{[ \text{انحدار } S_1 , S_2 - ( \text{انحدار } S_2 ) ]^2 / (2 - 1)}$$

مع ملاحظة أن  $22 - (ص) = 22 - ( \text{انحدار } S_1 , S_2 )$  هو الاختلاف غير المفسر . وبالمثل ، إذا أوجدنا الاختلاف الذي يفسره  $S_2$  وحده وطرحناه من الاختلاف الذي يفسره المتغيران معا نحصل على المساهمة الخاصة بالمتغير  $S_1$  بعد استبعاد أثر  $S_2$  ويمكن اختبار هذه المساهمة بنفس الطريقة .

ففى المثال - انظر الحل بغير الحاسب - وجدنا ما يلى :

$$٢٢ (ص) = \text{الاختلاف الكلى} = ١٤٤$$

$$٢٢ (انحدار س١ ، س٢) = \text{الاختلاف المفسر للمتغيرين معا} = ١٤٠,٣٥١٢$$

$$\text{الاختلاف غير المفسر} = ٣,٦٥ \text{ بدرجات حرية } ٣ - ٢$$

نحسب الآن الاختلاف الذى يفسره كل من  $س١$  ،  $س٢$  عندما يستعمل كل منهما على حدة .

$$\frac{٢٢ (ص) - ٢٢ (انحدار س١)}{٢٢ (س١)} = \frac{٢٢ (ص) - ٢٢ (س١)}{٢٢ (س١)}$$

$$١٢٥ = \frac{٢ \left( \frac{٤٠ \times ٢٥٠}{١٠} - ١٠٥٠ \right)}{٢٠} \text{ بدرجة واحدة من الحرية}$$

$$٢٢ (انحدار س٢) = \frac{٢ \left( \frac{١٢٠ \times ٢٥٠}{١٠} - ٣٠٤٠ \right)}{٢٤} = ٦٦,٦٧ \text{ بدرجة واحدة من الحرية}$$

من الصيغة (١٣) ، لاختبار الفرض  $\beta = ٠$

$$ف١ = \frac{٧٣,٦٨١٢}{٠,٥٢١٤} = \frac{٦٦,٦٧ - ١٤٠,٣٥١٢}{٧ \div ٣,٦٥}$$

$$١٤١,٣١٤ = \text{بدرجتى حرية } ١ ، ٧$$

ولاختبار الفرض  $\beta = ٠$

$$ف٢ = \frac{١٥,٣٥١٢}{٠,٥٢١٤} = \frac{١٢٥ - ١٤٠,٣٥١٢}{٧ \div ٣,٦٥}$$

$$٢٩,٤٤٢ = \text{بدرجتى حرية } ١ ، ٧$$



ونظراً لأن  $F_{[7, 17, 0.01]} = 12.2$  نرفض كلا من  $H_0: \beta = 0$  ،  $H_0: \beta = 0$  .

عند مستوى الدلالة 0.01 .

يلاحظ أن اختبار ف المستخدم هنا يكافئ اختبار ت الذي نتج عن الحل السابق ، ويتأكد هذا من ملاحظة أن  $\sqrt{141,314} = 11,888 = t_{17}$  ،  $\sqrt{29,442} = 5,426 = t_{17}$  . ويمكن إذا أردنا أن نلون هذه النتائج في جدول كالآتي .

الجدول (١٢ - ٦)

اختبار كل من المتغيرين بعد استبعاد أثر المتغير الآخر

مصدر الاختلاف	د. ح	ط. ح	ف
التحدر $S_1$ ، $S_2$ معا	٢		
التحدر $S_1$ وحده	١		
التحدر $S_2$ وحده	١		
التحدر $S_1$ بعد $S_2$	١	٧٣,٦٨	١٤١,٣١٤
التحدر $S_2$ بعد $S_1$	١	١٥,٣٥	٢٩,٢٤٢
الاختلاف غير المقسّر	٧	٠,٥٢١٤	
الكل	٩		

## ملاحظات :

من التحليل الملخص بالجدول (١٢ - ٦) نخرج بالتائج والملاحظات الآتية :

(١) المتغير  $S_1$  حين يستخدم وحده للتنبؤ بقيم  $S_2$  أفضل من المتغير  $S_3$  حين يستخدم وحده لنفس الغرض ، وذلك لأن الاختلاف الذى يفسره  $S_1$  منفردا وهو ١٢٥ أكبر من الاختلاف الذى يفسره  $S_3$  منفردا وهو ٦٦,٦٧ . ونصل إلى نفس النتيجة إذا قارنا معامل التحديد ، فمعامل التحديد للمتغير  $S_1$  هو :

$$r_{S_1}^2 = \frac{125}{144} = 0,86806 \text{ بينما } r_{S_3}^2 = \frac{66,67}{144} = 0,46299$$

وبهذا الأسلوب نستطيع مقارنة أى عدد من المتغيرات تستخدم فرادا .

(٢) فى تفسير الاختلاف الكلى فى  $S_2$  تكون مساهمة أى متغير أكبر فى حالة استخدامه منفردا عنها فى حالة استخدامه بعد متغير آخر فبالنسبة للمتغير  $S_1$  نجد أن  $125 < 73,68$  وبالنسبة للمتغير  $S_3$  نجد أن  $66,67 < 15,35$  وهذه نتيجة عامة مهما كان عدد المتغيرات التنبؤية ، فالمساهمة المنفردة لأى متغير تكون أكبر دائما من مساهمته مع متغير أو عدة متغيرات أخرى .

(٣) معامل التحديد المحسوب من معادلة الانحدار للمتغيرين  $S_1$  ،  $S_3$  معا وهو  $r_{(1,3)}^2 = \frac{140,35}{144} = 0,975$  أكبر من معامل التحديد المحسوب

من معادلة الانحدار لأى من المتغيرين وهما ٠,٨٦٨ ، ٠,٤٦٣ . وهذه نتيجة عامة ، فمعامل التحديد يقل دائما حين يستبعد واحد أو أكثر من المتغيرات التنبؤية .

(٤) فى معادلة الانحدار للمتغيرين  $S_1$  ،  $S_3$  معا يصعب تقدير الأهمية النسبية لهذين المتغيرين من حيث مقدار المساهمة فى التنبؤ بقيم المتغير التابع  $S_2$  ،

لأننا إذا قدرنا هذه الأهمية النسبية بواسطة الاختلافين المفسرين عند استخدام كل متغير على حدة وهما ١٢٥ ، ٦٦،٦٧ نجد أن هذا التقدير غير مناسب لأن مجموع هذين الاختلافين وهو ١٩١،٦٧ يزيد عن الاختلاف الكلي في ص وهو ١٤٤ . ومن ناحية أخرى ، إذا اعتبرنا الاختلاف الناشئ عن المتغير س<sub>١</sub> بعد استبعاد أثر المتغير س<sub>٢</sub> وهو ٧٣،٦٨ والاختلاف الناشئ عن المتغير س<sub>٢</sub> بعد استبعاد أثر المتغير س<sub>١</sub> وهو ١٥،٣٥ نجد أن مجموعهما وهو ٨٩،٠٣ يقل عن الاختلاف الذى يفسره المتغيران معا وهو ١٤٠،٣٥ . وهذه الصعوبة في تقدير الأهمية النسبية للمتغيرات هى إحدى الصعوبات التى يلقاها الباحث عند التصدى لمقارنة المتغيرات فى الانحدار الخطى المتعدد وعند اختيار أفضل المتغيرات التى تدخل فى معادلة الانحدار كما سيأتى بالبند التالى .

(٥) قيمة أى معامل انحدار جزئى ب<sub>١</sub> لمتغير س<sub>١</sub> هى قيمة شرطية تختلف باختلاف المتغيرات التى تدخل معه فى معادلة الانحدار .

إن أسلوب اختبار دلالة معاملات الانحدار الجزئية الذى أدى إلى الصيغة (١٣) هو أسلوب عام مهما كان عدد المتغيرات التنبؤية ، ويستخدم فى التعرف على دلالة ما يحدث من نقص فى دقة التنبؤ حين يستبعد واحد أو أكثر من المتغيرات من معادلة الانحدار . وبصفة عامة إذا كنا قد وفقنا معادلة انحدار فى ك من المتغيرات س<sub>١</sub> ، س<sub>٢</sub> ، ... ، س<sub>ك</sub> ووفقنا معادلة انحدار فى ل > ك من هذه المتغيرات أى بعد استبعاد ك - ل منها ، وأردنا اختبار ما إذا كانت دقة التنبؤ قد نقصت نقصا ذا دلالة نتيجة لهذا الاستبعاد فإننا نستخدم الصيغة العامة الآتية :

$$F = \frac{[ (ك) ٢ ٢ - (ل) ٢ ٢ ] / (ك - ل)}{[ (ص) ٢ ٢ - (ك) ٢ ٢ ] / (١ - ك - ل - ل)} \quad (١٤)$$

بدرجتي حرية ك - ل ، ل - ك - ١ ،

حيث ٢٢ (ك) هو الاختلاف الناشئ عن الانحدار على ك من المتغيرات  
 ، ٢٢ (ل) هو الاختلاف الناشئ عن الانحدار على ل من هذه المتغيرات  
 ، ٢٢ (ص) هو الاختلاف في قيم المتغير التابع ص  
 إذ يمكن إثبات أن كلا من بسط ومقام هذه النسبة هو تقدير مستقل للتباين  
 .<sup>٢٥</sup>

إن الصيغة (١٤) يمكن كتابتها بدلالة معاملات التحديد كالآتي :

$$(١٥) \quad F = \frac{[١ - (ك) - (ل)] / [(ل) - (ك)]}{[١ - (ك) - (ل)] / [(ك) - ١]}$$

بدرجتي حرية (ك - ل) ، (ل - ك - ١)

مثال (١٢ - ٢) :

أجرى انحدار خطي لمتغير عشوائي ص على خمسة متغيرات تنبؤية ص<sub>١</sub> ،  
 ص<sub>٢</sub> ، ص<sub>٣</sub> ، ص<sub>٤</sub> ، ص<sub>٥</sub> ، ووجد أن معامل التحديد في عينة حجمها ٤٦ هو  
 ٠,٦٢ وعندما أجرى انحدار خطي لنفس المتغير على ثلاثة من هذه المتغيرات  
 ص<sub>١</sub> ، ص<sub>٢</sub> ، ص<sub>٣</sub> وجد أن معامل التحديد ٠,٥٧ هل يمكن الاستغناء عن  
 المتغيرين ص<sub>٤</sub> ، ص<sub>٥</sub> والاكتفاء بالمتغيرات الثلاثة الأخرى دون أن نفقد شيئاً من  
 دقة التنبؤ بقيم المتغير ص ؟

الحل :

من الصيغة (١٥) :

$$F = \frac{٢ \div (٠,٥٧ - ٠,٦٢)}{٤٠ \div (٠,٦٢ - ١)} = ٢,٦٣٢$$

وهذه القيمة تقل عن القيمة الحرجة ف  $3,23 = [2,0, 4,0]$  مما يدعونا إلى قبول  
الفرض الصفري أن دقة التنبؤ لم تتأثر ، وبالتالي نستطيع الاكتفاء بالمتغيرات  
 $\beta_1, \beta_2, \beta_3$  للتنبؤ بالمتغير  $\beta$  دون أن نفقد شيئا من دقة التنبؤ .

مثال (١٢ - ٣) :

في المثال (١٢ - ١) هل نستطيع الاكتفاء بأحد المتغيرين للتنبؤ بقيم المتغير التابع  
 $\beta$  دون أن نفقد شيئا من دقة التنبؤ ؟

الحل :

سبق أن أجبنا عن هذا السؤال حين استخدمنا الصيغة (١٣) - التي هي حالة  
خاصة من الصيغة (١٤) أو (١٥) - في رفض كل من الفرضين  $\beta_1 = 0, \beta_2 = 0$   
لأن هذا يعنى أن وجود أى من المتغيرين له أثر ذو دلالة في عملية التنبؤ . باستخدام  
الصيغة (١٥) لاختبار إمكانية الاستغناء عن المتغير  $\beta_3$  لدينا :

$$\begin{aligned} \beta_3 &= 0,97466 \text{ ، } \beta_3 &= 0,46299 \\ \text{في } F &= \frac{1 \div (0,46299 - 0,97466)}{7 \div (0,97466 - 0,46299)} = \frac{0,51167}{0,00362} \end{aligned}$$

$$= 141,35$$

وهذه القيمة ذات دلالة عالية وتعنى أن الاستغناء عن  $\beta_3$  يقلل من دقة التنبؤ .  
وبالمثل ، لاختبار إمكانية الاستغناء عن  $\beta_1$

$$F = \frac{0,86806 - 0,97466}{0,00362} = 29,45 \text{ وهذه أيضا ذات دلالة عالية وتعنى}$$

أن الاستغناء عن  $\beta_1$  يقلل من دقة التنبؤ .

## (١٢ - ٤) اختيار متغيرات التنبؤ :

من الصعوبات التي يلقاها الباحثون في الانحدار المتعدد كيفية اختيار المتغيرات التي تدخل في معادلة الانحدار لتعطي أعلى درجة من الدقة في التنبؤ بالمتغير التابع ص . وفي المعتاد يكون لدى الباحث مجموعة كبيرة من المتغيرات المرشحة لذلك ويكون من المفيد عمليا اختصار هذه المجموعة إلى مجموعة تتكون من أقل عدد ممكن من هذه المتغيرات بشرط أن تعطي هذه المجموعة الجزئية معادلة تنبؤ لا تقل كفاءة عن المعادلة التي تستخدم جميع المتغيرات المتاحة . إن عملية الاختصار هذه ممكنة في الغالب لأن بعض المتغيرات المتاحة لا تسهم بدرجة كافية في عملية التنبؤ كما قد يبدو لأول وهلة ، كما أنه قد توجد مجموعة ( أو أكثر ) من المتغيرات المتاحة ترتبط ببعضها ارتباطا عاليا وبالتالي تعطي معلومات مماثلة ويمكن حينئذ الاكتفاء بمتغير واحد فقط من هذه المجموعة ليُمثلها جميعا .

ويتلخص السؤال المطروح هنا فيما يلي : إذا كان لدينا ك من المتغيرات التنبؤية المرشحة للدخول في معادلة الانحدار فكيف نختار أصغر مجموعة جزئية من هذه المجموعة بحيث تعطي نفس الدرجة من الدقة في التنبؤ بالمتغير التابع ؟ فمثلا إذا كان لدينا مجموعة من عشرة متغيرات تنبؤية فهل يمكن أن نكتفي بمجموعة من اثنين أو ثلاثة منها لتعبر عن المجموعة بكاملها ( دون أن نفقد شيئا من دقة التنبؤ ) ؟

الواقع أنه لا توجد إجابة واحدة شافية لهذا السؤال ولكن هناك عدة اقتراحات بمحاولات نسترشد بها في تحديد المجموعة المطلوبة ، وتتطلب هذه المحاولات النظر إلى البيانات المشاهدة عدة مرات من جوانب مختلفة ، وهنا يلعب الحاسب دورا رئيسيا لأن تنفيذ هذه المحاولات يحتاج إلى جهد حسابي شاق .

نفرض أننا حصلنا على بيانات مشاهدة في عينة حجمها  $n$  كما في الجدول (١٢ - ١) السابق .

(١) إن أول ما يخطر بالبال هو اختيار المتغيرات الأكثر علاقة بالمتغير التابع ص ، فنقوم بالمقارنة بينها كمتغيرات فرادا إما عن طريق مقادير الاختلافات المفسرة ٢٢ ( الانحدار ) أو عن طريق معاملات التحديد ص كما جاء بالملاحظة (١) بالبند (١٢ - ٣) السابق ، فتكون المتغيرات ذوات القيم الأكبر مرشحة مبدئيا للدخول في معادلة الانحدار متعدد .

(ب) نبحت عما إذا كان هناك مجموعة ( أو أكثر ) من المتغيرات التنبؤية التي ترتبط ببعضها ارتباطا عاليا ، فإذا وجدت مثل هذه المجموعة فإن أحدها يمكن أن يمثلها جميعا . ويحتاج الأمر هنا إلى إيجاد معامل الارتباط بين كل زوج من هذه المتغيرات ، ويفضل وضع هذه المعاملات وأيضا معاملات الارتباط بين كل من المتغيرات والمتغير التابع في جدول لتسهيل الرجوع إليها . وعلى فرض أن هناك خمس متغيرات تنبؤية يكون الجدول كالاتي :

١ ص	٢ ص	٣ ص	٤ ص	٥ ص
١ ص	٢١ ص	٣١ ص	٤١ ص	٥١ ص
٢ ص	١	٣٢ ص	٤٢ ص	٥٢ ص
٣ ص		١	٤٣ ص	٥٣ ص
٤ ص			١	٥٤ ص
٥ ص	١ ص	٢ ص	٣ ص	٤ ص

(ج) نوجد معادلة الانحدار ص على جميع ما لدينا من متغيرات ثم نفحص دلالة كل من معاملات الانحدار الجزئية ب ، ب ، ب ، ب ، ب عن طريق اختبار

ت بالصيغة (١٠) فتكون المتغيرات التي ترشح للدخول في معادلة الانحدار هي تلك التي تغطي بالقيم الأكبر من قيم ت ، أما المتغيرات التي تناظر القيم الأصغر من ت فتكون معرضة للاستبعاد . على أن ذلك لا ينبغي أن يتم دون تدبر فإن استخدام أو استبعاد متغير مالا يقرر لمجرد أن قيمة ت المناظرة كبيرة أو صغيرة ، لأن قيم معاملات الانحدار الجزئية وبالتالي قيم ت هي قيم شرطية وتتغير بتغير المتغيرات التي تدخل معها في الانحدار كما سبق القول بالملاحظة (٥) ، ومن ناحية أخرى قد توجد أسباب نظرية أو خبرات تحتم الاحتفاظ ببعض المتغيرات حتى ولو كان معامل الانحدار الجزئي المناظر غير ذي دلالة . فمثلا إذا كان لدينا عدة متغيرات نرغب استخدامها في التنبؤ بالمسافة التي تقطعها سيارة ما وكان من بين هذه المتغيرات متغير « مقدار الوقود المستهلك » فلا بد من الاحتفاظ بهذا المتغير حتى ولو كان معامل الانحدار المناظر له غير ذي دلالة .

(د) على ضوء ما نجده في (أ) و(ب) و(ج) نختار مجموعة أو أكثر من المتغيرات ونحسب معادلات انحدار كل منها مع إيجاد معامل التحديد ومقارنته بمعامل التحديد الذي ينجم من الانحدار على جميع المتغيرات المتاحة واختبار دلالة الفرق بينهما في دقة التنبؤ باستخدام الصيغة (١٥) . وإذا وجدنا أن إحدى المجموعات المختارة بنفس كفاءة المجموعة الكاملة نحاول اختزالها بنفس الطريقة إلى مجموعة ذات عدد أقل من المتغيرات .

وبيني اختيارنا النهائي لمجموعة المتغيرات التي تدخل في معادلة التنبؤ التي ننشدها على أساس أنها ( أولا ) تشتمل على أقل عدد من المتغيرات و( ثانيا ) يمكنها أن تعبر عن مجموعة المتغيرات المتاحة بكاملها ، أى بحيث لا تقل دقة التنبؤ منها عن دقة التنبؤ من استخدام جميع المتغيرات المتاحة . هذا ويجدر الإشارة أن المجموعة التي تختار على هذا الأساس ليست فريدة ، بل يمكن أن نعر على أكثر من مجموعة تستوفي الشرطين المطلوبين .



## تقارین (۱۲ - ۱)

أجريت دراسة لمعرفة أثر العوامل الجغرافية على حجم نوع من الضفادع واستهدف جزء من هذه الدراسة التنبؤ بطول جسم الضفدعة عن طريق الارتفاع عن سطح البحر والمتوسط السنوي لدرجة الحرارة . وقد جمعت بيانات من ۱۳ موقعا ودونت بالجدول الآتي :

الجدول (۱۲ - ۷)

الموقع	طول الضفدعة عن	الارتفاع عن سطح البحر	درجة الحرارة س <sup>۲</sup>
(۱)	۲۳,۷	۶۰۰	۶۰
(۲)	۲۳,۹	۸۰۰	۵۵
(۳)	۲۳,۴	۱۴۵۰	۵۰
(۴)	۲۳,۴	۱۴۰۰	۵۰
(۵)	۲۲,۲	۱۰۰	۵۵
(۶)	۱۹,۵	۴۵۰	۵۰
(۷)	۲۳,۵	۴۰۰	۵۰
(۸)	۲۴,۷	۸۰۰	۶۰
(۹)	۲۳,۳	۸۰۰	۶۰
(۱۰)	۲۱,۷	۱۱۰۰	۵۰
(۱۱)	۲۱,۳	۵۰۰	۵۰
(۱۲)	۲۰,۲	۴۰۰	۵۰
(۱۳)	۲۲,۱	۴۰۰	۵۰

$$\begin{aligned} 8490.00 &= \text{م ص}^2 = 92.00 = \text{م ص}^2 & 1127.27 &= \text{م ص}^2 = 292.9 = \text{م ص}^2 \\ 4860.00 &= \text{م ص}^2 = 3680.00 = \text{م ص}^2 & & & 69.0 = \text{م ص}^2 \\ 10092.0 &= \text{م ص}^2 = 21.265 = \text{م ص}^2 \end{aligned}$$

(أولا) استخدم مخرجات الحاسب الالكتروني المبينة بالجدول (١٢ - ٨) الآتي لإيجاد معادلة الانحدار الخطي للمتغير صـ على المتغيرين مـ<sub>١</sub> ، مـ<sub>٢</sub> .  
 (ثانيا) اختبر دلالة الانحدار ككل ودلالة كل من معامل الانحدار . هل من الممكن الاستغناء عن أحد المتغيرين التنبؤيين دون أن نفقد شيئا من دقة التنبؤ ؟  
 (ثالثا) أجب عن الجزئين (أولا) و(ثانيا) دون الاستعانة بمخرجات الحاسب .

الجدول (١٢ - ٨)

Regress. of	Y	Length of Frog		
On	X <sub>1</sub>	Altitude		
	X <sub>2</sub>	Temperature		
Variable Name	Regress. Coeff.	S.E. of Coeff.	t	D.F.
Constant	9.785579	4.129044	2.37	
X <sub>1</sub>	.1700912 E-02.	.81158094-03	2.10	10
X <sub>2</sub>	.2174479	.7589070 E-1	2.87	10
Multiple Corr. Coeff. (R) = .73379				
Coeff. of Determination (R <sup>2</sup> ) = .5384				
Estimated Standard Error of Estimate = 1.1390588				
Analysis of Variance for Regression				
Source of Variation	SS	D.F.	M S	F
Regression	15.13315	2	7.566575	5.8319
Residual	12.97455	10	1.297455	

ملاحظة :

الرمز 01 - E يعنى ضرب العدد الذى على يسار هذا الرمز فى  $10^{-1}$   
الرمز 02 - E يعنى ضرب العدد الذى على يسار هذا الرمز فى  $10^{-2}$   
وهكذا ...  
الرمز 02 + E يعنى ضرب العدد الذى على يسار هذا الرمز فى  $10^{+2}$   
وهكذا ...  
فمثلا العدد 02 - E 1700912 . يعنى العدد 001700912 .  
هذا ما يسمى بالرمزية العلمية scientific notation

## ثانيا - الارتباط الخطى المتعدد

سنفترض هنا أن لدينا  $k + 1$  من المتغيرات العشوائية  $x_1, x_2, \dots, x_k, x_{k+1}$  كما سنفترض أن توزيع الاحتمال المشترك لهذه المتغيرات هو توزيع معتدل متعدد المتغيرات multivariate normal distribution إن هذا الفرض يتضمن أمرين : الأول أن توزيع أى من هذه المتغيرات ( على حدة ) هو توزيع معتدل ، والثانى أن العلاقة بين أى من هذه المتغيرات وأى متغير آخر أو مجموعة من المتغيرات الأخرى هي علاقة خطية .

وامتدادا لدراساتنا فى الارتباط الخطى البسيط فى الفصل العاشر نتناول فى البندين الآتيين معاملين رئيسيين هما معامل الارتباط الخطى المتعدد ومعامل الارتباط الجزئى ونرى كيف نختبر دلالة كل منهما .

### ( ١٢ - ٥ ) معامل الارتباط المتعدد

#### MULTIPLE CORRELATION COEFFICIENT

إذا أردنا تقدير درجة العلاقة الخطية بين أحد هذه المتغيرات وليكن  $x_1$  وبقيّة المتغيرات ، فإننا نعرّف معامل الارتباط المتعدد بين  $x_1$  والمتغيرات الأخرى بأنه

معامل الارتباط البسيط بين المتغير  $x_1$  والمتغير العشوائى الذى يتكون من التركيب الخطى  $\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$  وسنرمز لهذا المعامل بالرمز  $r_{(1 \dots k)}$ .

ولتقدير هذا المعامل نأخذ عينة عشوائية من  $n$  من وحدات المجتمع ونقوم بقياس قيمة كل من هذه المتغيرات لكل وحدة فنحصل على  $n$  من المشاهدات كل منها على الصورة  $(x_{11}, x_{21}, \dots, x_{k1}, y_1)$  حيث  $y_1 = \alpha + \beta_1 x_{11} + \beta_2 x_{21} + \dots + \beta_k x_{k1} + \epsilon_1$  وتوضع هذه القياسات عادة كما فى الجدول الآتى .

جدول (١٢ - أ)

الملاحظات	$x_{11}$	$x_{21}$	$x_{k1}$	$y_1$	$\dots$	$x_{1n}$	$x_{2n}$	$x_{kn}$	$y_n$
(١)	$x_{11}$	$x_{21}$	$x_{k1}$	$y_1$	$\dots$	$x_{1n}$	$x_{2n}$	$x_{kn}$	$y_n$
(٢)	$x_{12}$	$x_{22}$	$x_{k2}$	$y_2$	$\dots$	$x_{1n}$	$x_{2n}$	$x_{kn}$	$y_n$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
(٣)	$x_{1r}$	$x_{2r}$	$x_{kr}$	$y_r$	$\dots$	$x_{1n}$	$x_{2n}$	$x_{kn}$	$y_n$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
(٥)	$x_{1n}$	$x_{2n}$	$x_{kn}$	$y_n$	$\dots$	$x_{1n}$	$x_{2n}$	$x_{kn}$	$y_n$

من هذه المشاهدات نحصل على التقدير المطلوب بعدد سنرمز له بالرمز  $r_{(1 \dots k)}$  ونحسبه كما فى حالة الارتباط الخطى البسيط من الصيغة

$$(١٦) \quad r_{(1 \dots k)}^2 = \frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلى}}$$

أى أن التقدير المطلوب هو الجذر التربيعى لنسبة الاختلاف الذى يفسره الانحدار الخطى المتعدد إلى الاختلاف الكلى فى المتغير  $x_j$  . ويمكن إثبات أنه فى هذه الحالة .

$$(17) \quad 1 \leq r_j \leq \sqrt{(1 + \dots + r_j^2)} \leq 0$$

أى أن هذا المعامل لا يمكن أن يكون سالبا وهو يختلف فى هذه الصفة عن معامل الارتباط البسيط الذى نعلم أنه يمكن أن يأخذ قيما سالبة .

وكما فى الانحدار الخطى المتعدد تعتمد فى حساب هذا المعامل على الحاسب الالكترونى توفيراً للجهد والوقت . وفى الحالة البسيطة التى يكون لدينا فيها ثلاثة متغيرات عشوائية  $x_1$  ،  $x_2$  ،  $x_3$  يمكننا حساب معامل الارتباط المتعدد  $r_{123}$  بين المتغير  $x_1$  والمتغيرين  $x_2$  ،  $x_3$  من الصيغة الآتية :

$$(18) \quad \frac{r_{12}r_{13} - r_{23}r_{11} - r_{11}r_{23} + r_{11}r_{22}}{r_{22}r_{33} - 1} = r_{1(23)}^2$$

ويتطلب حساب هذه القيمة إيجاد كل من :

- $r_{11}$  وهو معامل الارتباط البسيط بين المتغيرين  $x_1$  ،  $x_1$  ،
- $r_{12}$  ، وهو معامل الارتباط البسيط بين المتغيرين  $x_1$  ،  $x_2$  ،
- $r_{13}$  ، وهو معامل الارتباط البسيط بين المتغيرين  $x_1$  ،  $x_3$  ،

ومن الواضح أن قيمة معامل الارتباط المتعدد تعتمد على قيم معاملات الارتباط البسيط بين أزواج المتغيرات المستخدمة ، وهذه الملاحظة صحيحة مهما كان عدد المتغيرات .

## اختبار دلالة معامل الارتباط المتعدد :

لاختبار الفرض الصفري  $H_0 = (1 + e \dots 32) = 0$  ضد الفرض  $H_1 = (1 + e \dots 32) \neq 0$  نستخدم نفس الصيغة (٦) التي وردت بالبند (١٢ - ١) وهي :

$$F = \frac{\text{التباين المفسر}}{\text{التباين غير المفسر}}$$

$$(19) \quad \frac{e / (1 + e \dots 32) 1^2}{(1 - e - v) / (1 + e \dots 32) 1^2 - 1} =$$

إذا توفرت الشروط المذكور في مستهل الجزء الثاني من هذا الفصل فإن هذه الإحصاءة يكون لها توزيع ف بدرجتي حرية  $e$  ،  $v - e - 1$  حيث  $e + 1$  عدد جميع المتغيرات المستخدمة . يلاحظ أن كلا من الصيغتين (٦) ، (١٩) تختبر وجود أو عدم وجود العلاقة الخطية .

مثال (١٢ - ٤) :

في عينة عشوائية من ٢٣ مجموعة من القيم من مجتمع معتدل ذي ثلاثة متغيرات عشوائية وجدت معاملات الارتباط البسيطة الآتية :  $r_{11} = 0,38$  ،  $r_{12} = 0,46$  ،  $r_{22} = 0,51$  . أوجد معامل الارتباط المتعدد  $r_{(32)1}$  واختبر دلالة .

الحل :

$$\begin{aligned} & \text{من الصيغة (١٨) :} \\ & \frac{0,51 \times 0,46 \times 0,38 \times 2 - 0,2116 + 0,1444}{0,2601 - 1} = \frac{0,2116 + 0,1444}{0,2601 - 1} = 0,240 = \end{aligned}$$

∴ معامل الارتباط المتعدد المطلوب  $= \sqrt{0,240} = 0,49$  تقريبا .

من الصيغة (١٩) :

$$F = \frac{2 \div 0,240}{20 \div (0,24 - 1)} = 3,1079 \quad \text{بلرجنى حرية } 20,2$$

وهذه القيمة ليست بذات دلالة عند المستوى ٠,٠٥ وتشير إلى عدم وجود ارتباط خطى بين المتغير  $x_1$  والمتغيرين  $x_2$  ،  $x_3$  .

(١٢ - ٦) معامل الارتباط الجزئى

### PARTIAL CORRELATION COEFFICIENT

إن معامل الارتباط البسيط  $r_{12}$  بين متغيرين عشوائيين  $x_1$  ،  $x_2$  يفترض أن هذين المتغيرين يتأثران ببعضهما فقط ولا يتأثران بأى متغير آخر . إلا أنه فى كثير من الحالات يكون كل من هذين المتغيرين متأثرا تأثرا خطيا بمتغير عشوائى ثالث  $x_3$  ، وفى هذه الحالة يقتضى القياس الصحيح للارتباط بين  $x_1$  ،  $x_2$  استبعاد أثر هذا المتغير الثالث من كل منهما . إن المعامل الذى يقيس مثل هذا الارتباط يسمى بمعامل الارتباط الجزئى بين  $x_1$  ،  $x_2$  وسنرمز له بالرمز  $r_{12.3}$  . ويعرف هذا المعامل أيضا بأنه معامل الارتباط البسيط بين  $x_1$  ،  $x_2$  عند قيمة ثابتة للمتغير  $x_3$  أى فى قطاع الوحدات التى تأخذ نفس القيمة للمتغير  $x_3$  . ويتفق هذان التعريفان طالما كان توزيع الاحتمال المشترك للمتغيرات الثلاثة معتدلا  $trivariate \text{ normal distribution}$  إذ أنه فى هذه الحالة تظل قيمة  $r_{12.3}$  كما هى مهما كانت القيمة الثابتة للمتغير  $x_3$  ، وهذه هى الحالة التى نتناولها هنا والتى تتمشى مع ما افترضناه فى مستهل هذا الجزء .

ويقدر هذا المعامل من العينة بمقدار سنرمز له بالرمز  $r_{12.3}$  حيث

$$(20) \quad r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

وحيث  $r_{12}$  ،  $r_{13}$  ،  $r_{23}$  هي معاملات الارتباط البسيط بين أزواج المتغيرات .  
وبنفس الطريقة نحسب كلا من  $r_{12}$  ،  $r_{13}$  مع تذكر أننا هنا لا نميز بين  
متغير مستقل ومتغير تابع .

أما دلالة هذا المعامل فتختبره عن طريق الإحصاءة :

$$(21) \quad \text{بدرجات حرية } n - 3 \quad t = \frac{r_{12} \sqrt{n-3}}{\sqrt{1-r_{12}^2}}$$

أو عن طريق الإحصاءة المكافئة:

$$(22) \quad \text{بدرجات حرية } 1, n - 3 \quad F = \frac{r_{12}^2}{(1-r_{12}^2) / (n-3)}$$

مثال (١٢ - ٥) :

في دراسة للعلاقة بين ضغط الدم وكلوسترول الدم أخذت عينة عشوائية من  
١٤٢ من السيدات كبار السن وقيس كل من ضغط الدم  $x_1$  ودرجة تركيز  
الكلوسترول  $x_2$  والعمر  $x_3$  ، ثم حسبت معاملات الارتباط البسيط لكل زوج  
من هذه المتغيرات فوجدت كما يلي :  $r_{12} = 0.2495$  ،  $r_{13} = 0.3332$  ،  
 $r_{23} = 0.5029$  . بحث دلالة معامل الارتباط بين ضغط الدم ودرجة تركيز  
الكلوسترول (أولاً) دون استبعاد متغير العمر ، (ثانياً) بعد استبعاد متغير  
العمر .

الحل :

(أولاً) معامل الارتباط بين ضغط الدم ودرجة تركيز الكلوسترول دون



استبعاد متغير العمر هو  $r_{11} = 0.2495$  ونختبر دلالة هذا المعامل باستخدام الصيغة (٨) بالبند (١٠-٥) وهي :

$$t = \frac{r_{11} \sqrt{2-1}}{\sqrt{r_{11}^2 - 1}} = \text{بدرجات حرية } 2-1$$

$$\text{أو الصيغة المكافئة} = \frac{r_{11}^2}{(2-1)/(r_{11}^2 - 1)} = \text{بدرجات حرية } 1, 2-1$$

$$\text{لدينا : } t = \frac{\sqrt{140} \cdot 0.2495}{\sqrt{(0.2495)^2 - 1}} = 3.0486 = \text{بدرجات حرية } 140$$

وهذه القيمة تزيد عن القيمة الحرجة  $t_{0.01, 140} = 2.576$  فهي ذات دلالة عالية وتشير إلى وجود ارتباط خطي بين المتغيرين .

(ثانيا ) معامل الارتباط بين ضغط الدم ودرجة تركيز الكلسترول بعد استبعاد متغير العمر هو معامل الارتباط الجزئي  $r_{3-21}$  . من الصيغة (٢٠) نجد أن

$$r_{3-21} = \frac{0.5029 \times 0.3332 - 0.2495}{\sqrt{(0.5029^2 - 1)(0.3332^2 - 1)}} = 0.1233$$

ونختبر دلالة هذا المعامل من الصيغة (٢١) كالآتي :

$$t = \frac{0.1233 \sqrt{139}}{\sqrt{0.1233^2 - 1}} = 1.465 = \text{بدرجات حرية } 139$$

وهذه القيمة تقل عن القيمة الحرجة  $t_{(1198), 0.05} = 1.98$  فهي ليست بذات دلالة عند المستوى 0.05 ، وتشير إلى أنه حين نستبعد عامل العمر لا نستطيع الاستدلال على وجود ارتباط خطي بين ضغط الدم وكلوسترول الدم . هذا مع ملاحظة أن كلا من ضغط الدم ومقدار الكلوسترول في الدم يزداد بزيادة العمر ، ومن هنا نرى أن النتيجة التي حصلنا عليها في (أولا) كانت نتيجة مضللة .

### معاملات الارتباط الجزئي من مراتب أعلى :

إن معامل الارتباط الجزئي من النوع  $r_{xy.z}$  يوصف بأنه معامل ارتباط جزئي من المرتبة الأولى of the first order لأنه يعطى معامل الارتباط بين متغيرين بعد استبعاد أثر متغير واحد من كل منهما . وبالمثل إذا كان لدينا أربعة متغيرات عشوائية  $x_1, x_2, x_3, x_4$  ، فإننا نرمز بالرمز  $r_{x_1x_2x_3x_4}$  لمعامل الارتباط الجزئي بين المتغيرين  $x_1, x_2$  بعد استبعاد أثر المتغيرين  $x_3, x_4$  من كل منهما ، ونصف هذا المعامل حيثئذ بأنه معامل ارتباط جزئي من المرتبة الثانية . وبحسب هذا المعامل من معاملات الارتباط الجزئية من المرتبة الأولى كالآتي :

$$(23) \quad \frac{r_{x_1x_2} - r_{x_1x_3}r_{x_2x_3}}{(1 - r_{x_1x_3}^2)(1 - r_{x_2x_3}^2)} = r_{x_1x_2x_3}$$

أو كالآتي :

$$(24) \quad \frac{r_{x_1x_2} - r_{x_1x_4}r_{x_2x_4}}{(1 - r_{x_1x_4}^2)(1 - r_{x_2x_4}^2)} = r_{x_1x_2x_4}$$

ونختبر دلالة هذا المعامل من الصيغة :

$$T = \frac{\sqrt{4 - 5} \sqrt{43 - 21}}{\sqrt{43 - 21 - 1}} =$$

بدرجات حرية ٥ - ٤ (٢٥)

ويمتد مفهوم معاملات الارتباط الجزئية لأي عدد منتهى  $k$  من المتغيرات العشوائية التي تشترك في توزيع معتدل متعدد المتغيرات . ومعامل الارتباط الجزئي من المرتبة ٢ ( حيث  $2 < k$  ) هو معامل الارتباط بين اثنين من المتغيرات بعد استبعاد أثر ٢ من المتغيرات العشوائية الأخرى . ويعتمد حساب معامل الارتباط الجزئي من المرتبة ٢ على معاملات الارتباط الجزئية من المرتبة السابقة عليها أى من المرتبة ٢ - ١ . هذا ويجدر بنا ملاحظة التماثل في صيغ هذه المعاملات . أما اختبار دلالة معامل الارتباط الجزئي من المرتبة ٢ فهو تعميم للصيغتين (٢١) ، (٢٥) ويأخذ الصيغة الآتية :

$$T = \frac{\sqrt{2 - 2 - 5} \sqrt{3 - 2 - 5}}{\sqrt{3 - 2 - 1}} =$$

بدرجات حرية ٥ - ٢ - ٢ (٢٦)

حيث  $s$  هو معامل الارتباط الجزئي من المرتبة ٢ .

## تمارين (١٢ - ٢)

(١) في عينة عشوائية من ٣٠ وحدة من مجتمع معتدل ذي ثلاثة متغيرات وجد أن معامل الارتباط المتعدد  $s_{(32)} = ٠,٥$  . بين أن هذه القيمة تدل على وجود ارتباط في المجتمع بين المتغير  $x_1$  والمتغيرين  $x_2$  ،  $x_3$  مستخدما مستوى الدلالة ٠,٠٥

(٢) في عينة عشوائية حجمها ٣٩ من مجتمع معتدل ذي ثلاثة متغيرات وجد أن معاملات الارتباط البسيط بين أزواج هذه المتغيرات هي  $r_{12} = 0,28$  ،  $r_{13} = 0,49$  ،  $r_{23} = 0,51$  أثبت أن  $r_{12} = 0,445$  ،  $r_{13} = 0,42$  واختبر دلالة كل منهما .

(٣) بين أن معامل الارتباط الجزئي من المرتبة الثانية والذي قيمته  $r_{12.3} = 0,5$  المحسوب من عينة عشوائية حجمها ٢٠ من مجتمع معتدل ذي أربعة متغيرات يكون ذا دلالة عند المستوى  $0,05$

(٤) في عينة عشوائية من ٥٤ من السيدات حسبت المقادير المستهلكة في فترة ما من نوعين من الغذاء هما البروتين (ب) والدهون (د) كما حسب العمر ص. وقد وجدت معاملات الارتباط البسيط الآتية :

$r_{12} = 0,5784$  ،  $r_{13} = -0,4865$  ،  $r_{23} = -0,5296$  أوجد معامل الارتباط الجزئي بين استهلاك البروتين واستهلاك الدهون مستقلاً عن أثر العمر ، واختبر دلالة عند المستوى  $0,01$

## الفصل الثالث عشر

### دالة التمييز

#### DISCRIMINANT FUNCTION

يتناول هذا الفصل المشكلة الآتية . نفرض أن باحثا يبحث بمجمعين  $A$  ،  $B$  يشتركان بدرجات متفاوتة في بعض الخواص التي يمكن قياسها عدديا . إذا كان لدى الباحث عينة عشوائية يعلم أنها من أحد هذين المجمعين ولكنه لا يعلم ما إذا كانت من المجتمع  $A$  أو من المجتمع  $B$  ، وإذا كان الباحث قد قام بقياس وحدات هذه العينة من حيث  $K$  من تلك الخواص وحصل على القيم العددية  $x_1, x_2, \dots, x_K$  ، فكيف يستخدم هذه القيم لتحديد المجتمع الذي تنتمي إليه العينة ؟

فمثلا إذا قامت إحدى شركات التنقيب عن البترول بحفر بئر ووجدت منطقة رملية على عمق ٤٠٠٠ قدما فكيف تقرر عن طريق قياس بعض الخواص الجيوفيزيائية لعينة مأخوذة من هذه المنطقة ما إذا كانت المنطقة تحتزن بترولاً (المجتمع  $A$ ) فتستمر في الحفر ، أو لا تحتزن بترولاً (المجتمع  $B$ ) فتتوقف عن الحفر ؟ كذلك إذا كانت إحدى الكليات تقوم بفحص الطلاب المتقدمين إليها فكيف تميز بين الطلاب الذين يصلحون للدراسة فيها (المجتمع  $A$ ) والطلاب الذين لا يصلحون لذلك (المجتمع  $B$ ) عن طريق إجراء بعض الاختبارات العلمية والنفسية على الطلاب ؟

(١٣ - ١) دالة التمييز :

إن مثل هذه المشكلات تحل عن طريق إيجاد دالة  $D(x_1, x_2, \dots, x_K)$  ،

س) في المتغيرات التي تعبر عن تلك الخواص ، وعدد د. يقسم قيم هذه الدالة إلى جزئين بحيث إذا كانت قيمة هذه الدالة عند القياسات المشاهدة في عينة ما أصغر من العدد د. تكون العينة من المجتمع أ وإذا كانت أكبر من أو تساوى العدد د. تكون العينة من المجتمع ب ، مع بيان احتمال الخطأ في هذا التقسيم . وتسمى هذه الدالة حيثخذ بدالة التمييز كما يسمى العدد د. بالنقطة الحدية boundary point .

فمثلا في حفر بحر البترول قد تكون دالة التمييز هي :

د (س<sub>١</sub> ، س<sub>٢</sub> ، ... ، س<sub>٢٦</sub>) = س<sub>٢</sub> - س<sub>١</sub> - س<sub>٣</sub> + س<sub>٤</sub> - س<sub>٥</sub> + س<sub>٦</sub> ، ... ، س<sub>٢٦</sub> .  
والعدد د. = ٢٦ بحيث إذا كانت قيمة هذه الدالة لقياسات أخذت على وحدات عينة ما أصغر من ٢٦ يستمر الحفر ( المجتمع أ ) وإذا كانت أكبر من أو تساوى ٢٦ يتوقف الحفر ( المجتمع ب ) . نفرض مثلاً أن القيم المشاهدة للخواص الخمسة في إحدى العينات هي على الترتيب ٣٠ ، ١٢٠ ، ١٠ ، ١٦ ، ٢٠٠ . بالتعويض بهذه القيم في دالة التمييز نجد أن د (٣٠ ، ١٢٠ ، ١٠ ، ١٦ ، ٢٠٠) = ٣٨ وبما أن ٣٨ أكبر من ٢٦ تكون العينة من المجتمع ب ويتوقف الحفر .

### (١٣ - ٢) إيجاد دالة التمييز :

كيف نعد الدالة د والعدد د. لتكون جاهزة للتطبيق على عينات يراد معرفة المجتمع الذي تنتمي إليه ؟ الطريق إلى ذلك ما يلي :

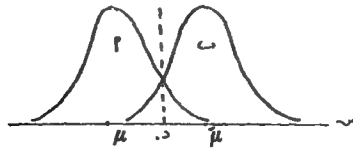
نأخذ عينة عشوائية حجمها ن<sub>١</sub> من المجتمع الأول وعينة عشوائية حجمها ن<sub>٢</sub> من المجتمع الثاني ، ونحدد ك من الخواص التي نرى أنها ذات تأثير في التمييز بين المجتمعين (ك > ن<sub>١</sub> ، ك > ن<sub>٢</sub>) ثم نقيس كل وحدة من وحدات العينة من حيث هذه الخواص فتكون القياسات الناتجة وعددها (ن<sub>١</sub> + ن<sub>٢</sub>) ك هي الأساس الذي نبني عليه دالة التمييز كما سيتبين بعد . فمثلاً إذا كانت ك = ٤ تكون البيانات كما في الجدول (١٣ - ١) الآتي :

المجلد (١٣ - ١)

عينة المجتمع الأول					عينة المجتمع الثاني				
س <sub>١</sub>	س <sub>٢</sub>	س <sub>٣</sub>	س <sub>٤</sub>	س <sub>٥</sub>	س <sub>١</sub>	س <sub>٢</sub>	س <sub>٣</sub>	س <sub>٤</sub>	س <sub>٥</sub>
١١	٢١	٣١	٤١	١	١١	٢١	٣١	٤١	١
١٢	٢٢	٣٢	٤٢	٢	١٢	٢٢	٣٢	٤٢	٢
٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠
٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠
١,٥	٢,٥	٣,٥	٤,٥	٢	١,٥	٢,٥	٣,٥	٤,٥	٢
متوسط	س <sub>١</sub>	س <sub>٢</sub>	س <sub>٣</sub>	س <sub>٤</sub>	متوسط	س <sub>١</sub>	س <sub>٢</sub>	س <sub>٣</sub>	س <sub>٤</sub>

(أولاً) : حالة متغير واحد :

لتقديم الفكرة التي تؤسس عليها دالة التمييز ، نبدأ بالحالة التي يكون لدينا فيها متغير واحد  $\mu$  (ك = ١) . سنفترض أن لهذا المتغير توزيعاً محدداً بمتوسط  $\mu$  للمجتمع الأول ،  $\mu'$  للمجتمع الثاني ، أما التباين فقيمته واحدة للمجتمعين وقدرها  $\sigma^2$  . يمثل هذان التوزيعان كما في الشكل (١٣ - ١) الآتي .



الشكل (١٣ - ١)

نفرض أن  $s$  هي قيمة المتغير  $s$  التي وجدناها في عينة نعلم أنها من أحد المجتمعين ونريد تحديد المجتمع الذي تنتمي إليه العينة . إن دالة التمييز هنا تكون على الصورة  $d(s) = s$  . من الطبيعي أن نأخذ متوسط المجتمعين  $d = \frac{1}{2}(\mu + \mu')$  كنقطة حدية تستخدم للفصل بينهما . وإذا فرضنا أن  $\mu$  أصغر من  $\mu'$  فإن قاعدة التمييز تكون كالآتي :

إذا كان  $s > \frac{1}{2}(\mu + \mu')$  نقرر أن العينة من المجتمع أ

وإذا كان  $s \leq \frac{1}{2}(\mu + \mu')$  نقرر أن العينة من المجتمع ب  
ونتساءل الآن متى يكون قرارنا خاطئاً ؟ إن هذا الخطأ يقع حين تكون العينة في الواقع من المجتمع أ ويكون قرارنا أنها من المجتمع ب وهذا يحدث حيناً  
( مع أن العينة من المجتمع أ )  $s \leq \frac{1}{2}(\mu + \mu')$

$$\text{أى حيناً } \frac{\mu - \mu'}{\sigma \sqrt{2}} = \frac{\mu - (\frac{1}{2}(\mu + \mu'))}{\sigma} \leq \frac{\mu - s}{\sigma}$$

$$\text{أى حيناً } \frac{\delta}{\sigma \sqrt{2}} \leq \epsilon$$

$$(1) \quad \mu - \mu' = \delta , \quad \frac{\mu - s}{\sigma} = \epsilon \quad \text{حيث}$$

$$(2) \quad \text{واحتيال هذا الخطأ هو } (\frac{\delta}{\sigma \sqrt{2}} \leq \epsilon)$$

ويسمى هذا الاحتمال باحتيال خطأ التقسيم . ونحصل على نفس قيمة هذا الاحتمال حين تكون العينة من المجتمع ب ونقرر أنها من المجتمع أ . تحقق من ذلك .



ولما كانت  $\epsilon$  تتبع التوزيع المعتدل القياسي فإننا نستطيع إيجاد هذا الاحتمال من جدول المساحات أسفل المنحنى المعتدل القياسي .

ويلاحظ أنه إذا كان المتوسطان  $\mu$  ،  $\mu'$  متساويين أو متساويين تقريبا فإن  $\frac{\delta}{\sigma}$  تكون تقريبا مساوية للصفر ويكون احتمال خطأ التقسيم مساويا بالتقريب

للنسبة ٥٠٪. ومعنى هذا أن التقسيم عشوائيا وفي هذه الحالة لا يكون هناك جدوى من إيجاد دالة صالحة للتمييز بين المجتمعين . كما يلاحظ أن احتمال خطأ التقسيم يكون أصغر ما يمكن إذا كانت  $\left| \frac{\mu - \mu'}{\sigma} \right| = \left| \frac{\delta}{\sigma} \right|$  أكبر ما يمكن أى إذا كانت

القيمة الموجبة لنسبة الفرق بين متوسطى المجتمعين إلى الانحراف المياري المشترك أكبر ما يمكن ، وتستخدم هذه الحقيقة كأساس لاشتقاق أفضل دالة للتمييز بين المجتمعين .

( ثانيا ) : حالة  $k$  من المتغيرات :

نأتى الآن للحالة التى يكون لدينا فيها  $k$  من المتغيرات  $x_1, x_2, \dots, x_k$  ، كما سبق القول ، للتوصل إلى دالة التمييز نأخذ عينة عشوائية من كلا المجتمعين ونقيس قيم هذه المتغيرات لكل وحدة من وحدات العينتين فتكون البيانات الناتجة هى الأساس الذى نبني عليه إنشاء أفضل دالة تمييز كما يتبين بعد .

إمكانية التحليل الاحصائى سنضع الافتراضات الثلاثة الآتية .

افتراضات التحليل :

(١) دالة التمييز خطية أى على الصورة

$$D(x) = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k$$

$$(٢) \quad \sum_{i=1}^k \alpha_i^2 = 1$$

حيث  $\alpha_1, \alpha_2, \dots, \alpha_k$  بارامترات مجهولة مطلوب تقديرها من العيتين بشرط أن نحصل من هذه التقديرات على أقل احتمال لخطأ التقسيم .

(٢) المتغيرات  $x_1, x_2, \dots, x_k$  هي متغيرات عشوائية توزيعها المشترك هو توزيع معتدل متعدد ( ذك من المتغيرات ) سواء في المجتمع  $\alpha$  أو في المجتمع  $\beta$  . وتأخذ متوسطات هذه المتغيرات القيم  $\mu_1, \mu_2, \dots, \mu_k$  في المجتمع الأول ، والقيم  $\mu'_1, \mu'_2, \dots, \mu'_k$  في المجتمع الثاني . أما التباين  $\sigma^2$  لأي متغير  $x_i$  والذي سنرمز له بالرمز  $\sigma_i^2$  فهو واحد في المجتمعين وتختلف التباينات من متغير إلى آخر ، كما أن التباين  $\sigma_i^2$  ( $i \neq j$ ) لمتغيرين  $x_i, x_j$  يكون واحدا في المجتمعين وتختلف التباينات بين كل زوج وآخر من المتغيرات .

من هذا الافتراض ينتج أن دالة التمييز المعرفة في (٣) تكون ذات توزيع معتدل لأنها خطية في متغيرات معتدلة . هذا مع تذكر أن اعتدالية التوزيع المتعدد المشترك تتضمن اعتدالية كل متغير على حدة .

(٣) العينات التي تؤخذ من كل مجتمع هي عينات عشوائية .

لايجاد أفضل تقدير لدالة التمييز المعرفة في (٣) من أزواج العينات ينبغي أن نقدر البارامترات  $\alpha_1, \alpha_2, \dots, \alpha_k$  بأعداد  $a_1, a_2, \dots, a_k$  تجعل الدالة  $L$  ذات أقل خطأ ممكن للتقسيم . وهذا الشرط كما جاء في حالة المتغير الواحد يكافئ أن تكون النسبة  $\left| \frac{\delta}{\sigma} \right|$  أكبر ما يمكن حيث  $\delta, \sigma$  هنا تعرفان كالآتي :

$\delta$  = الفرق الكلي بين المتوسطات المتناظرة في المجتمعين مرجحة بالمعاملات  $\alpha_1, \alpha_2, \dots, \alpha_k$

$$= \sum_{i=1}^k \alpha_i (\mu_i - \mu'_i)$$

$$= \sum_{i=1}^k \alpha_i \delta_i \quad \text{حيث } \delta_i = \mu_i - \mu'_i \quad (٤)$$

،  $\sigma' =$  التباين الكلي للدالة (٣) وهي  $\alpha \alpha'$

$$(٥) \quad \sigma' = \sum_{i=1}^k \alpha_i \alpha_i' = \sum_{i=1}^k \sigma_i^2$$

فمثلا إذا كانت  $k = ٣$  فإن

$$\delta = (\bar{\mu}_1 - \mu_1) \alpha_1 + (\bar{\mu}_2 - \mu_2) \alpha_2 + (\bar{\mu}_3 - \mu_3) \alpha_3$$

$$\sigma' = \alpha_1' \alpha_1 \sigma_1^2 + \alpha_2' \alpha_2 \sigma_2^2 + \alpha_3' \alpha_3 \sigma_3^2$$

$$+ \alpha_4' \alpha_4 \sigma_4^2 + \alpha_5' \alpha_5 \sigma_5^2 + \alpha_6' \alpha_6 \sigma_6^2$$

المطلوب إذن إيجاد التقديرات  $\alpha_i$  بحيث تكون القيمة المطلقة  $\left| \frac{\delta}{\sigma} \right| = \Delta$

أكبر ما يمكن حيث  $\delta$ ،  $\sigma$  معرفتان في (٤)، (٥). ولتجنب الإشارات سنأخذ  $\Delta' = \frac{\delta}{\sigma}$  بدلا من  $\Delta$ . أى أننا سنقدر البارامترات بالقيم التى تجعل الدالة

الآتية أكبر ما يمكن :

$$\Delta' = \frac{\delta}{\sigma} = \frac{\alpha_1 \alpha_1' + \alpha_2 \alpha_2' + \alpha_3 \alpha_3' + \alpha_4 \alpha_4' + \alpha_5 \alpha_5' + \alpha_6 \alpha_6'}{\alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + \alpha_3^2 \sigma_3^2 + \alpha_4^2 \sigma_4^2 + \alpha_5^2 \sigma_5^2 + \alpha_6^2 \sigma_6^2}$$

$$(٦) \quad \Delta' = \frac{\alpha_1 \alpha_1' + \alpha_2 \alpha_2' + \alpha_3 \alpha_3' + \alpha_4 \alpha_4' + \alpha_5 \alpha_5' + \alpha_6 \alpha_6'}{\alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + \alpha_3^2 \sigma_3^2 + \alpha_4^2 \sigma_4^2 + \alpha_5^2 \sigma_5^2 + \alpha_6^2 \sigma_6^2}$$

حيث  $\alpha_r =$  مجموع المربعات للمتغير  $x_r$  في المجتمعين ( $r = 1, 2, \dots$ ) (ك)

$\alpha_r =$  مجموع حواصل الضرب للمتغيرين  $x_r$  ،  $x_s$  في المجتمعين .  
وتعظيم الدالة (٦) يكافئ تعظيم الدالة

$$(٧) \quad \delta = \alpha_1^2 + \alpha_2^2 + \dots + \alpha_r^2 + \alpha_{r+1}^2 + \dots + \alpha_s^2$$

التي لا تختلف عنها إلا في المقدار الثابت  $\alpha_1^2 + \alpha_2^2 + \dots + \alpha_r^2$  . ويتقضى هذا التعظيم إجراء التفاضل الجزئي لهذه الدالة بالنسبة إلى كل من  $\alpha_1$  ،  $\alpha_2$  ، ... ،  $\alpha_s$  ، والمساواة بالصفر . وهذا يؤدي إلى الحصول على  $s$  من المعادلات الخطية في  $s$  من المجاهيل تشبه المعادلات المعتادة في الانحدار الخطي ، فهي تأخذ الصورة الآتية :

$$\delta = \alpha_1^2 + \alpha_2^2 + \dots + \alpha_r^2 + \alpha_{r+1}^2 + \dots + \alpha_s^2$$

$$\delta = \alpha_1^2 + \alpha_2^2 + \dots + \alpha_r^2 + \alpha_{r+1}^2 + \dots + \alpha_s^2$$

$$(٨) \quad \dots \dots \dots \dots \dots \dots$$

$$\delta = \alpha_1^2 + \alpha_2^2 + \dots + \alpha_r^2 + \alpha_{r+1}^2 + \dots + \alpha_s^2$$

والقيم  $\alpha_1$  ،  $\alpha_2$  ، ... ،  $\alpha_s$  التي تحقق هذه المعادلات جميعا تكون هي التقديرات المطلوبة للبارامترات  $\alpha_1$  ،  $\alpha_2$  ، ... ،  $\alpha_s$  وبالتالي تكون دالة التمييز هي :

$$(٩) \quad \alpha_1^2 + \alpha_2^2 + \dots + \alpha_r^2 + \alpha_{r+1}^2 + \dots + \alpha_s^2$$

هذا مع ملاحظة أن مجاميع المربعات ومجاميع حواصل الضرب تقدر من العيتين كالآتي :

$\Sigma x^2 =$  مجموع المربعات لقيم المتغير  $x$  في العينة الأولى + مجموع المربعات لقيم نفس المتغير في العينة الثانية .

وبالمثل لمجاميع المربعات الأخرى  $\Sigma y^2$  ،  $\Sigma z^2$  ،  $\Sigma w^2$  ،  $\Sigma v^2$  = مجموع حواصل الضرب للمتغيرين  $x$  ،  $y$  في العينة الأولى + مجموع حواصل الضرب لنفس المتغيرين في العينة الثانية

وبالمثل لمجاميع حواصل الضرب الأخرى  $\Sigma xy$  ،  $\Sigma xz$  ،  $\Sigma xv$  ،  $\Sigma xw$  ،  $\Sigma yz$  ،  $\Sigma yv$  ،  $\Sigma yw$  ،  $\Sigma zv$  ،  $\Sigma zw$  ،  $\Sigma vw$  .

ولإيجاد هذه القيم ثم حل المعادلات (٨) يحتاج إلى الحاسب الالكتروني توفيراً للجهد وضماناً للدقة والسرعة .

### النقطة الحدية :

كما في حالة المتغير الواحد ، من الطبيعي أن نأخذ متوسط المجتمعين كنقطة حدية تفصل بينهما ، وعلى ذلك تقدر النقطة الحدية  $D$  من العيتين كالآتي :

$$D = \frac{1}{2} \left( \bar{x}_1 + \bar{x}_2 \right) \quad (١٠)$$

وهكذا نكون قد حصلنا على دالة التمييز (٩) والقيمة الحدية (١٠) . فإذا حصلنا على قياسات  $x_1$  ،  $x_2$  ،  $x_3$  ،  $x_4$  ،  $x_5$  لعينة جديدة ، يمكن بالتعويض بهذه القياسات في الدالة (٩) ثم المقارنة بالعدد  $D$  أن نقرر ما إذا كانت العينة تنتمي إلى المجتمع أ أو إلى المجتمع ب .

## احتمال خطأ التقسيم PROBABILITY OF MISCLASSIFICATION

يبقى أن نقدر احتمال خطأ التقسيم لدالة التمييز التي أوجدناها . وكما في حالة المتغير الواحد تكون أكبر قيمة لهذا الاحتمال هي المعطاة بالصيغة (٢) وهي :

$$(١١) \quad \delta \leq \epsilon \left( \frac{\delta}{\sigma^2} \right)$$

حيث  $\delta$  ،  $\sigma$  تعطيان بالصيغتين (٤) ، (٥) ، ويمكن إثبات أن  $\frac{\delta}{\sigma}$  تقدر من العينتين

كما يلي :

$$(١٢) \quad \frac{\delta}{\sigma} = \sqrt{\frac{1}{2} (v_1 + v_2 - 2)} \quad \text{حيث } \bar{s}_1 - \bar{s}_2 = \frac{\delta}{\sigma}$$

،  $\bar{s}_1$  الوسط الحسابي للمتغير  $s_1$  في العينة الأولى

،  $\bar{s}_2$  الوسط الحسابي للمتغير  $s_2$  في العينة الثانية .

والاحتمال (١١) يمكن إيجاده من جدول المساحات أسفل المنحنى المعتدل القياسي وهو يشير إلى درجة الثقة في دالة التمييز ، ويعتبر التقسيم جيدا بدرجة عالية إذا لم يزد هذا الاحتمال عن ٧٪ .

ملاحظات :

(١) إذا كانت دالة التمييز هي د (  $s_1$  ،  $s_2$  ، ... ،  $s_p$  ) والقيمة الحدية د. ، وحسبنا من أزواج العينات القيمتين د (  $\bar{s}_1$  ،  $\bar{s}_2$  ، ... ،  $\bar{s}_p$  ) ،

د (س<sub>١</sub> ، س<sub>٢</sub> ، ... ، س<sub>١٠</sub>) فينبغي أن تكون إحدى هاتين القيمتين أكبر من د .  
والأخرى أصغر منها .

(٢) إذا كان كل زوج من متوسطات المجتمعين متساويين (م<sub>١</sub> = م<sub>٢</sub>) أو متساويين تقريبا فلا أمل في العثور على دالة تميز بين المجتمعين بكفاءة . ولهذا نجب العناية باختيار المتغيرات التي تستخدم في دوال التمييز بحيث تكون هناك فروق معقولة بين متوسطاتها .

(٣) استخدام الحاسب الآلي في هذه الدراسة أمر ضروري لسرعة ودقة ما نريد التوصل إليه ، خاصة وأنها تضطر في كثير من الحالات إلى تجربة مجموعات مختلفة من المتغيرات لاختيار الأصلح منها . وهذا يتطلب مشقة كبيرة يغبينا عنها الحاسب .

مثال (١٣ - ١) :

في دراسة لتوزيع بكتريا النيتروجين Azotobacter في التربة كان المطلوب معرفة مدى دقة التنبؤ بوجود أو عدم وجود هذه البكتريا في التربة ، أو بمعنى آخر المطلوب إيجاد دالة خطية تميز التربة التي تحتوي على هذه البكتريا من التربة التي لا تحتوي عليها ، وذلك باستخدام ثلاث خواص كيميائية للتربة هي :

م<sub>١</sub> = الأس الأيدروجيني pH ، م<sub>٢</sub> = كمية الفوسفات الموجودة ، م<sub>٣</sub> = المحتوى الكلي للنيتروجين . وقد أخذت عينة عشوائية حجمها ن<sub>١</sub> = ١٠٠ من تربة لا تحتوي على هذه البكتريا وعينة عشوائية حجمها ن<sub>٢</sub> = ١٨٦ من تربة تحتوي عليها . وقسمت كل وحدة من وحدات العيتين من حيث هذه الخواص الثلاث وحسبت الفروق ف<sub>١</sub> = س<sub>١</sub> - س<sub>٢</sub> بين أزواج المتوسطات في العيتين فوجدت كما يلي :

$$ف_١ = ٠,١٤٠٨ ، ف_٢ = ٠,٠٨٢١ ، ف_٣ = ٠,٠٨٢٦$$

وحسبت مجاميع المربعات  $\chi^2$  لكل من المتغيرات الثلاثة ، ومجاميع حواصل الضرب  $\chi^2$  لأزواج المتغيرات فوجدت كما يلي :

$$\chi^2_{11} = 1,111 , \chi^2_{22} = 1,043 , \chi^2_{33} = 2,942$$

$$\chi^2_{12} = 0,229 , \chi^2_{13} = 0,198 , \chi^2_{23} = 0,001$$

وبذلك كانت المعادلات المعتادة كالآتي :

$$1,111 + \alpha_1 0,229 + \alpha_2 0,198 + \alpha_3 0,1408 = 0$$

$$0,229 + \alpha_1 0,229 + \alpha_2 1,043 + \alpha_3 0,0821 = 0$$

$$0,198 + \alpha_1 0,229 + \alpha_2 0,001 + \alpha_3 2,942 = 0$$

أستخدم الحاسب الالكتروني في حل هذه المعادلات فأتتج القيم الآتية :

$$\alpha_1 = 0,11229 , \alpha_2 = 0,00310 , \alpha_3 = 0,1960 \text{ وبذلك كانت دالة التمييز هي :}$$

$$D(\alpha_1, \alpha_2, \alpha_3) = 0,11229 \alpha_1 + 0,00310 \alpha_2 + 0,1960 \alpha_3$$

ويمكن حساب النقطة الحدية  $D$  من الصيغة (١٠) .

نقدر أكبر احتمال لخطأ التقسيم بهذه الدالة كما يلي :

$$\begin{aligned} \alpha_1 \text{ فـ} &= 0,11229 \times 0,1408 + 0,00310 \times 0,0821 + 0,1960 \times 0,08226 \\ &= 0,02179 \end{aligned}$$



$$\text{من (١٢) : قيمة } \frac{\delta}{\sigma} \text{ مقدرة من العينة} = \sqrt{0,02179 \times (2 - 186 + 100)}$$

$$2,4876 =$$

$$\text{من (١١) أكبر احتمال لخطأ التقسيم } J = \left( \frac{\delta}{\sigma} \leq \varepsilon \right)$$

$$J = (1,2438 \leq \varepsilon)$$

$$= 10,75\% \text{ (من جدول المساحات أسفل المنحنى المعتدل القياسي)}$$

قد يكون من المناسب مقارنة هذا الاحتمال باحتمال خطأ التقسيم الناشء من دالة تمييز استخدمت واحدا فقط من المتغيرات الثلاث .

نفرض أننا استخدمنا المتغير  $\mu$  فقط . لدينا :

$$\text{فم} = 0,1408 \text{ وهذه تقدير لقيمة } \delta$$

$$\text{الانحراف المعياري} = \sqrt{\frac{1}{2 - \frac{1}{2} + \frac{1}{2}}} = \sqrt{\frac{1,111}{284}} = 0,0625$$

وهذا تقدير  $\sigma$  .

$$\therefore J = \left( \frac{\delta}{\sigma} \leq \varepsilon \right) = \left( \frac{0,1408}{0,0625 \times 2} \leq \varepsilon \right) J = (1,126 \leq \varepsilon)$$

$$= 12,92\%$$

وإذا استخدمنا المتغير  $\delta$  فقط نجد بنفس الطريقة ما يلي :

$$J = \left( \frac{\delta}{\sigma^2} \leq \epsilon \right) \quad J = (0,58 < \epsilon) = 0,28,1 \%$$

أما إذا استخدمنا المتغير  $\delta$  فقط فإن

$$J = \left( \frac{\delta}{\sigma^2} \leq \epsilon \right) \quad J = (0,41 \leq \epsilon) = 0,34,09 \%$$

ويتضح أن أفضل دالة تمييز تؤسس على متغير واحد فقط هي تلك التي تستخدم المتغير  $\delta$  ، ولديها تلك التي تستخدم المتغير  $\delta$  ثم تلك التي تستخدم المتغير  $\delta$  وهذه الأخيرة تكون دالة ضعيفة للغاية لا يجوز الاعتماد عليها . على أن دالة التمييز المركبة من المتغيرات الثلاثة معا هي أفضلها جميعا .

### (١٣ - ٣) اختبار تساوى أزواج المتوسطات - اختبار ت<sup>٢</sup> .

كما سبق الذكر في الملاحظة (٧) بالبند السابق ينبغي أن تكون المتغيرات التي تستخدم في إعداد دالة التمييز هي تلك المتغيرات الأكثر قدرة على التمييز بين المجتمعين أى التي تختلف متوسطاتها في المجتمعين اختلافا معقولا . ولذلك يهتما في اختبار هذه المتغيرات أن نبحت دلالة الفروق بين أزواج المتوسطات في العينتين ، فإذا كانت هذه الفروق ليست ذات دلالة بمعنى أن أزواج المتوسطات في المجتمعين متساوية ، لا تكون دالة التمييز قادرة على فصل المجتمعين . أما إذا كانت تلك الفروق جوهرية فإن فرصة دالة التمييز في تقسيم المجتمعين بكفاءة تكون كبيرة . وهناك اختبار ابتكره هوتلينج Hotelling يسمى اختبار ت<sup>٢</sup> يمكننا من اختبار

الفرض الصفري المركب عن تساوي كل زوج من المتوسطات ، أى اختبار الفرض الصفري :

ملحوظة = ملحق لجميع  $r = 1, 2, \dots, k$  (١٣)  
ويعتمد هذا الاختبار على تحليل التباين للمتغير  $x$  المرتبط إلى مركبتين : « بين » و « داخل » المجتمعين كالآتي :

$$s^2 = (\text{داخل العينات}) = \frac{SS}{n-1} = \frac{SS}{n-1}$$

$$= \text{مجموع الحريات بدرجات حرية } n_1 + n_2 + \dots + n_k - k - 1 \quad (١٤)$$

(نتج الصيغة الأخيرة من ضرب المعادلات (٨) في  $\alpha_1, \alpha_2, \dots, \alpha_k$  ثم الجمع مع وضع  $x$  بدلا من  $\alpha$  ، فـ  $x$  بدلا من  $\alpha$  ) .

$$s^2 = (\text{بين العينات}) = \frac{SS}{n_1 + n_2 + \dots + n_k} = \text{مجموع الحريات } k \quad (١٥)$$

وقد أخذت  $k$  كدرجة الحرية بين العينات لأن التقديرات  $x$  اختبرت على أساس تعظيم النسبة بين  $s^2$  ( بين العينات ) ،  $s^2$  ( داخل العينات ) .

في المثال السابق نجد ما يلي :

$$s^2 = (\text{داخل العينات}) = \text{مجموع الحريات } k$$

$$= 0.02179 \quad (\text{سابق حسابه}) \quad \text{بدرجات حرية } 282$$

$$٢٢ \text{ (بين العينات) } = \frac{٢٧ + ١٧}{٢٧ + ١٧} = (ع ا ر ف)٢$$

$$٢(٠,٠٢١٧٩) \frac{١٨٦ \times ١٠٠}{٢٨٦} =$$

$$٠,٣٠٨٨ =$$

بدرجات حرية ٣

وينشأ لدينا جدول التباين الآتي .

المجدول (١٣ - ٧)

تحليل التباين لدالة التمييز - المتغيرات ٢

مصدر التباين	٢ ٢	درجات الحرية	ط ٢
بين نوعي التربة	$٠,٣٠٨٨ = (ع ا ر ف)٢$	$٣ = ك$	$٠,٠١٠٢٩$
داخل نوعي التربة	$٠,٠٢١٧٩ = ع ا ر ف$	$٢٧ + ١٧$	$٠,٠٠٠٠٧٧$
		$٢٨٢ = ١ - ك$	

$$١٣٣,١ = ٠,٠٠٠٠٧٧ \div ٠,٠١٠٢٩ = \text{في}^{**}$$

هذه القيمة أكبر بكثير من القيمة الحرجة ف  $٠,٠١$  [٢٨٢,٣] التي لا تزيد عن ٣,٧٨ فهي ذات دلالة عالية وتدعونا إلى رفض الفرض الصفري عن تساوى أزواج المتوسطات في المجتمعين ، وهذا ما يجب أن يكون إذا كانت دالة التمييز ذات كفاءة في فصل المجتمعين .

يلاحظ أنه إذا ظهر أن في غير ذات دلالة فإن دالة التمييز تفشل في فصل المجتمعين وينبغي حينئذ البحث عن متغيرات أخرى أو البحث عن دالة أخرى غير خطية فيما لدينا من متغيرات .

#### (١٣ - ٤) استخدامات دالة التمييز :

إن دالة التمييز هي وسيلة لدراسة مدى تداخل المجتمعات في بعضها أو مدى تباعدها عن بعضها . ولله الدالة ثلاثة أنواع من الاستخدامات تلخص فيما يلي :

##### (١) التقسيم والتشخيص :

يتبين هذا الهدف من المثال المقدم في بداية هذا الفصل حيث استخدمت دالة للتمييز بين المناطق التي تحتوي بترولاً والمناطق التي لا تحتوي عليه مستعينة بخمسة متغيرات جيوفيزيائية ، وكذلك من المثال (١٣ - ١) حيث أعدت دالة في ثلاثة متغيرات كيميائية تقسم التربة إلى نوعين يحتوي أحدهما على بكتريا النتروجين ولا يحتوي الآخر عليها . كذلك إذا كان هناك نوعان من الحمى يتشابهان في الأعراض فمن المفيد أن يكتشف الطبيب القياسات الجسمية والمعملية التي تساعد على التمييز بين نوعي الحمى وأن يعرف الطريقة المثل لضم هذه القياسات في دالة واحدة ، وكيفية تقدير درجة الثقة في تشخيص المرض .

##### (٢) دراسة العلاقات بين المجتمعات :

فمثلاً ، إلى أي مدى تختلف الاتجاهات والاستعدادات النفسية للمهندسين الكفاء عنها في رجال الأعمال الكفاء ؟ أو هل يختلف المدخنون عن غير المدخنين اختلافاً جوهرياً في السمات النفسية والعادات السلوكية ؟

##### (٣) تعميم لاختبارات :

إذا أجرينا عدة قياسات على كل من عيتين عشوائيتين أخذتا من مجتمعين معطلين ورغبنا في استخدام اختبار واحد للفرض الصفري عن تساوي متوسطات

هذين المجتمعين بالنسبة لجميع هذه القياسات فإن دالة التمييز تمكنتنا من ذلك عن طريق اختبارات<sup>٢</sup> كما جاء بالهند السابق .

هذا وتجدر الإشارة إلى أن الدراسة التي قدمت في هذا الفصل عن دالة التمييز تمتد إلى الحالات التي تتناول أكثر من مجتمعين ، كما تمتد إلى الحالات التي تكون فيها دالة التمييز غير خطية .

## الفصل الرابع عشر

### الطرق غير البارامترية

#### NONPARAMETRIC METHODS

إن معظم اختبارات الفروض التي جاءت في الفصول السابقة كانت تتطلب افتراض أن للمجتمع الذى نعين منه توزيعاً معتدلاً أو يمكن تقريبه بتوزيع معتدل ، كما أن بعضها كان يتطلب افتراضات أخرى مثل تساوى التباينات أو استقلال العينات . وجدير بالذكر أن هذه الاختبارات يمكن الاطمئنان إليها حتى لو وجدت انحرافات بسيطة عن هذه الافتراضات غير أن هناك مواقف يستحيل فيها قبول مثل هذه الافتراضات ولهذا كان من الضروري أن تنشأ أساليب أخرى تبني على افتراضات أقل صرامة . وقد عرفت هذه الأساليب بالطرق غير البارامترية أو بالطرق حرة التوزيع *distribution-free methods* وهى طرق يمكن تطبيقها على مدى واسع من التوزيعات دون أن تفترض توزيعاً محدداً لما تتناوله من مجتمعات . وتوصف هذه الطرق بأنها غير بارامترية لأن أغلبها لا يهتم باختبار أو تقدير بارامترات ( أدلة ) المجتمع .

وفضلاً عن أن الطرق والاختبارات غير البارامترية تستخدم تحت شروط عامة للغاية وتعطيناً من القلق عن صحة الافتراضات فهى تتميز بعدة أمور منها :

(١) أنها عادة ما تكون أسهل في الفهم والتفسير من تلك الطرق القياسية التي تعمل كبديل لها .

- (٢) أن ما تتطلبه من عمليات حسامية تكون عادة سهلة وسريعة .  
(٣) أنها لا تشترط أن تكون البيانات كمية ( عديدة ) بل يمكن أن تكون نوعية أو ترتيبية .

ولهذا شاع استخدام الطرق غير البارامترية بالرغم من أنها لا تعطي نفس القدر من المعلومات أو الدقة التي تعطيها الطرق البارامترية للمناظرة لها فهي بصفة عامة أقل كفاءة منها . وإذا وجد موقف يمكن فيه تطبيق كلا الأسلوبين فينبغي دائماً استخدام الأسلوب البارامترى فهو الأكثر كفاءة .

وقد مر بنا مثالان للاختبارات غير البارامترية جاء أحدهما بالبند (٦ - ٧) عند استخدام اختبار  $\chi^2$  ، وجاء الآخر في البندين (١٠ - ٧) و(١٠ - ٨) عند دراسة معامل ارتباط الرتب ( سيرمان ) ودلالته ، ونقدم فيما يلى ستة من الاختبارات الأخرى الشهيرة .

#### (١٤ - ١) اختبار العلاقات ( للكشف عن عشوائية العينة ) : RUNS TEST

إن جميع طرق الاستدلال الإحصائي التي نوقشت في الفصول السابقة بنيت على افتراض أن العينات عشوائية وعلى أن التجارب التي نحصل منها على البيانات صممت على هذا الأساس . غير أن هناك حالات يصعب فيها تحديد مدى تحقق هذا الافتراض وينبئ حيثئذ اختبار عشوائية العينة قبل التصدى لتحليل البيانات .

وتظهر هذه الحالات بصفة خاصة عندما نكون عاجزين كلياً أو جزئياً عن التحكم في اختيار العينة . فمثلاً في تقدير معدل الوفاة من مرض معين لا مفر من الاعتماد على سجلات سابقة وهذه لا تشكل عينة عشوائية بالمعنى الدقيق . كذلك الحال حين لا يكون لنا خيار إلا الاعتماد على أى سجلات متاحة لإعطاء تنبؤات عن الأحوال الجوية أو دراسة حوادث المرور أو حين نأخذ وحدات صناعية بحسب ترتيب إنتاجها .



ويتم اختيار عشوائية العينة الذي نقدمه هنا على نظرية تسمى بنظرية التلاحقات theory of runs التي تعتمد بدورها على الترتيب الذي سحبت به عناصر العينة . ويهدف الاختبار إلى الكشف عما إذا كان هذا الترتيب عشوائياً أو يتخذ نمطاً معيناً لا يمكن أن نعزوه إلى الصدفة .

اعتبر متتابعة تنقسم عناصرها إلى صنفين فقط . إن أى متتابعة جزئية تتألف من حد أقصى لعناصر متتالية من أحد هذين الصنفين تسمى تلاحقة . فمثلاً إذا اخترنا ١٢ شخصاً وكان الحرف «ك» يرمز إلى أن الشخص «ذكر» والحرف «ث» يرمز إلى أن الشخص أنثى فإن المتتابعة .

ك ك ك ك ك ك ك ك ك ك ك ك ك ك

تشتمل على ٥ تلاحقات ، تتألف الأولى من اثنين من الكافات ونقول إن طولها اثنان، وتتألف الثانية من ثلاثة من الثاعات ونقول إن طولها ثلاثة ، وتتألف الثالثة من كاف واحدة .. .. وهكذا .

وسواء كانت البيانات نوعية أو كمية فإن اختبار التلاحقات يقسمها إلى صنفين متنافيين : ذكور أو إناث - وحدة معينة أو غير معينة - مريض أو غير مريض - فوق الوسيط أو تحت الوسيط .. وهكذا .

ليكن  $n$  = حجم العينة ،  $s$  = عدد التلاحقات

$n_1$  = عدد المرات التي يظهر فيها الحرف الذي يرمز إلى أحد الصنفين .

$n_2$  = عدد المرات التي يظهر فيها الحرف الذي يرمز إلى الصنف

الآخر .

فمثلاً في متتابعة التلاحقات المذكورة : لدينا  $n = 12$  ،  $s = 5$  ،

$n_1 = 7$  ،  $n_2 = 5$

إن اختبار التلاحقات مؤسس على الفكرة الآتية . إذا كان بالعينة عدد قليل جداً من التلاحقات ، مثلاً :



التوزيع المعتدل هو توزيع لمتغير متصل فإن الصيغة (٣) تصحح كالاتي تعويضاً عن هذا الاختلاف .

$$(٤) \quad \frac{m - (\frac{1}{4} \pm m)}{n} = m$$

وتؤخذ العلامة السالبة إذا كان عدد التلاحقات (س) في العينة يزيد عن الوسط الحسابي  $m$  المحسوب من الصيغة (١) ، وتؤخذ العلامة الموجبة إذا كانت س تقل عن  $m$  . هذا ويمكن الاستغناء عن هذا التصحيح وإهمال النصف إذا كانت العينة كبيرة .

وحين تكون كل من  $n$  ،  $n$  أكبر أو تساوى عشرة نكون أمام واحد من الحالات الثلاثة الآتية :

(١) إذا كان الفرض الصفري هو أن العينة عشوائية والفرض الآخر هو العكس فإننا نستخدم اختباراً ذا جانبيين ونرفض الفرض الصفري عند مستوى الدلالة ٠,٠٥ إذا وقعت القيمة المشاهدة للإحصاءة (٤) خارج المنطقة :

$$(٥) \quad (١,٩٦ < m < -١,٩٦)$$

ونرفضه عند المستوى ٠,٠١ إذا وقعت خارج المنطقة :

$$(٦) \quad (٢,٥٨ < m < -٢,٥٨)$$

وهذا بحسب ما جاء بالمثال (٤ - ٣) في الفصل الرابع . ويمكن أن نوجد المناطق المناظرة لأى مستوى دلالة آخر من جدول المساحات أسفل التحتي المعتدل المعياري .

(ب) أما إذا كان عدد التلاحقات صغيراً وأردنا اختبار الفرض عن وجود نمط يفسر قلة هذه التلاحقات فإن الاختبار في هذه الحالة يكون ذا جانب واحد هو الجانب الأيسر ونرفض الفرض الصفري عن عشوائية العينة لصالح هذا الفرض عند مستوى الدلالة ٠,٠٥ إذا كانت القيمة المشاهدة للإحصاءة (٤) أقل من - ١,٦٤ ونرفضه عند مستوى الدلالة ٠,٠١ إذا كانت تقل عن - ٢,٣٣ . راجع المثال (٤ - ٣) .

(ج) كذلك إذا كان عدد التلاحقات كبيراً وأردنا اختبار الفرض عن وجود غلط دورى فإن الاختبار يكون أيضاً ذا جانب واحد هو الجانب الأيمن ونرفض الفرض الصفري عن عشوائية العينة لصالح هذا الفرض عند المستوى ٠,٠٥ (أو ٠,٠١) إذا كانت القيمة المشاهدة للإحصاءة (٤) تزيد عن ١,٦٤ (أو ٢,٣٣).

مثال (۱۴ - ۱) :

أُخذت عينة من ٣٠ شجرة دُرْدَاء كانت قد زرعت من عدة سنوات على طريق زراعي فوجدت المتابعة الآتية ، حيث ص تعبر عن أن الشجرة مصابة بمرض معين ، ح تعبر عن أنها غير مصابة . اختبر عند مستوى الدلالة ٠,٠١ ، الفرض بوجود تجمعات أى أن الأشجار المصابة تتجمع معاً .

ح ح ح ح ح ح ح ح    ص ص ص    ح ح ح ح ح

ح ح ح ح    ص ص ص ص ص    ح ح    ص ص

**الحل :**

اللدنيا  $s = 7$  (عدد التلاحقات) ،  $u = 20$  (عدد الأشجار السليمة) ،  
 $v = 10$  (عدد الأشجار المصابة) .

نظراً لأن كلا من  $\mu_1$  ،  $\mu_2$  لا تقل عن عشرة ، يمكن اعتبار توزيع الإحصاءة (4) معديلاً معيارياً .

$$\text{من (١) : } \mu = 1 + \frac{10 \times 20 \times 2}{10 + 20} = 14,23$$

$$\text{من (٢) : } \sigma = \sqrt{\frac{(10 - 20 - 10 \times 20 \times 2) \cdot 10 \times 20 \times 2}{(1 - 10 + 20) \times 30}} = 2,38$$

$$\text{بالتعويض في (٤) : } \text{ص} = \frac{14,23 - (\frac{1}{2} + 7)}{2,38} = 2,87$$

الفرض الصفري : العينة عشوائية .

الفرض الآخر : يوجد تجمعات . ( وإذن الاختبار ذو جانب واحد هو الجانب الأيسر )

بما أن  $2,87 > 2,33 -$  نرفض الفرض الصفري عند مستوى الدلالة ٠,٠١ ، عن عشوائية العينة ويكون لدينا دليل قوى على أن الأشجار المصابة تقع في تجمعات غير عشوائية .

### حالة البيانات الكمية .

لاختبار عشوائية العينة في حالة البيانات الكمية كأن يكون لدينا متتابعة تعبر عن أوزان حيوان ينمو سجلت في فترات يومية أو أسبوعية ، نستخدم نفس الأسلوب السابق إلا أننا في هذه الحالة نقسم ما لدينا من قيم إلى صنفين بحسب وقوعها فوق الوسيط أو تحت الوسيط فنضع حرفاً أ مثلاً ( أو علامة + ) لكل قيمة تزيد عن الوسيط وحرفاً ب ( أو العلامة - ) لكل قيمة تقل عن الوسيط مع الاحتفاظ بترتيب هذه القيم . وإذا وجدت قيم تساوى الوسيط فإنها تحمل وكأنها لم تكن .

( نذكر أنه لإيجاد الوسيط لمجموعة من الأعداد نرتب هذه الأعداد تصاعدياً أو تنازلياً ثم نأخذ العدد الذى فى الوسط إذا كان عدد هذه الأعداد فردياً ، أو متوسط العددين الأوسطين إذا كان عدد الأعداد زوجياً ) .

إن طريقة التلاحقات أعلى وأسفل الوسيط تفيد على وجه الخصوص فى حالتين رئيسيتين أولهما اختبار الاتجاهات وثانيهما اختبار الأنماط الدورية . فإذا بدأت متتابعة التلاحقات بحروف أغلبها ا ثم بحروف أغلبها ب فإن هناك انجهاً إلى أسفل ، وإذا بدأت بحروف أغلبها ب ثم بحروف أغلبها ا فإن هناك ميلاً إلى أعلى . أما إذا كان الحرفان ا ، ب يتبادلان بشكل منتظم فإن هذا يشير إلى وجود نمط دورى .

مثال ( ١٤ - ٢ ) :

أخذ قطاع على أرض متملحة ، وعند نقط محددة منه قدرت نسب الغطاء النباتى لنوع من النبات بعرض ٥ سم من القطاع . سجلت ٤٠ من هذه النسب بالترتيب كما يلى :

٢٧	٢٦	٣٥	٣٠	١٨	٢١	٢٨	٣٣	٢٧	٢٧	٢٤
١٩	٢٤	٢٦	٢٨	٢٨	٣١	١٩	٢٣	٣٢	٢٩	٢٤
٢٤	١٣	٢٣	٢١	١٧	٢٨	٢٣	٢٢	٢٣	٢٢	١٤
				١٧	١٠	٢١	٢١	١٨	٢٠	٢٨

اختبر ما إذا كان هناك نمط تجمع للغطاء النباتى .

الحل :

$$\text{الوسيط فى هذه العينة} = \frac{1}{2}(23 + 24) = 23,5$$

( يلاحظ أن إيجاد الوسيط يستلزم أولاً ترتيب الأعداد المعطاة تصاعدياً أو



### حالة العينات الصغيرة :

إذا كان أحد العددين  $n_1$  ،  $n_2$  أو كلاهما أصغر من ١٠ فلا يجوز التقريب بالتوزيع المحتل المعيارى . وفي هذه الحالة نستخدم جداول خاصة كالجدول (١٣) بملحق هذا الكتاب الذى يعطى احتمال أن تقل عدد التلاحقات عن عدد معين  $s$  ( على أساس صحة الفرض الصفرى عن عشوائية العينة ) عند زوج مرتب من الأعداد  $(n_1, n_2)$  أى يعطى الاحتمال :

ل  $(s \geq s_0)$  | ف صحيح عند  $(n_1, n_2)$  وحيث  $n_1 > n_2$

وقد أعد هذا الجدول بحيث يكون الإحداثي الأول  $n_1$  في الزوج المرتب  $(n_1, n_2)$  أصغر من الإحداثي الثاني  $n_2$  ، أى أننا نرمز بالرمز  $n_1$  لعدد مرات ظهور الحرف الذى يتكرر أقل سواء كان هذا الحرف هو  $a$  أو  $b$  . أما إذا كان  $n_1 = n_2$  فلا يوجد أى شرط . ونرفض الفرض الصفرى عن عشوائية العينة عند المستوى  $\alpha$  ( ٠,٥ مثلا ) إذا كان الاحتمال الناتج يقل عن  $\alpha$  وإلا نقبل الفرض الصفرى .

مثال (١٤ - ٣) :

ضبطت آلة لكى تصب مقداراً معيناً من سائل ما في كل وعاء يمر تحتها . وجد أنه في ١٥ وعاء متتالياً كانت مقادير السائل باللترات كالآتي :

٣,٦    ٣,٩    ٤,١    ٣,٦    ٣,٨    ٣,٧    ٣,٤    ٤,٠

٣,٨    ٤,١    ٣,٩    ٤,٠    ٣,٨    ٤,٢    ٤,١

هل نستطيع القول بأن المقادير التى توزعها الآلة تتغير عشوائياً ؟

الحل :

الوسيط = ٣,٩



بوضع ١ بدلا من كل عدد يزيد عن ٣,٩ ، ب بدلا من كل عدد يقل عن ٣,٩ وإهمال العددين المساويين للعدد ٣,٩ نحصل على المتابعة الآتية :

ب أ ب ب ب ب أ ب أ ب أ أ

لدينا س = ٨ (عدد التلاحقات) ، ن<sub>١</sub> = ٦ ، ن<sub>٢</sub> = ٧

مع ملاحظة أن ن<sub>١</sub> أصغر من ن<sub>٢</sub> وأن كلا منهما أصغر من ١٠

من الجدول (١٣) عند (ن<sub>١</sub> ، ن<sub>٢</sub>) = (٦ ، ٧) ، س = ٨ نجد أن :

ل (س ≥ ٨ | ف صحيح) = ٠,٧٣٣

وهذا الاحتمال يزيد عن ٠,٠٥ ولذلك نقبل الفرض الصفري أن حدود المتابعة تتغير عشوائياً .

(١٤ - ١ - ١) تطبيق آخر لاختبار التلاحقات :

يستخدم اختبار التلاحقات في الكشف عن دلالة تأثير معالجة ما على متغير ما كما يتبين من المثال الآتي .

مثال (١٤ - ٤) :

لمعرفة تأثير هورمون ما على أطوال براعم أحد النباتات أخذت عينة عشوائية من ٢٢ من هذا النبات وقسمت عشوائياً إلى قسمين بكل منهما ١١ نباتاً ووضعت النباتات تحت نفس الظروف فيما عدا أن نباتات أحد القسمين عولجت بالهورمون وتركزت نباتات القسم الآخر دون معالجة (مجموعة مراقبة) . وبعد أسبوعين وجد أن أطوال البراعم بالملليمترات كالآتي :

المجموعة المعالجة : ٢٦ ٢٨ ٣٧ ٤١ ٥٧ ٥٨ ٦٧ ٧٨ ٨٠ ١٢١ ١٢٢  
مجموعة المراقبة : ١٥ ٣٠ ٣٠ ٦٥ ٩١ ١٠٣ ١٣٢ ١٣٥ ١٣٧ ١٤١ ١٩٨  
ابحث ما إذا كان الهورمون يعيق نمو براعم هذا النبات .

الحل :

نضم جميع القيم المشاهدة في المجموعتين في متابعة واحدة ونكتب عناصر هذه المتابعة بعد ترتيبها ترتيبا تصاعديا ثم نضع الحرف ا تحت كل عنصر من عناصر المجموعة المعالجة والحرف ب تحت كل عنصر من عناصر مجموعة المراقبة كما يلي :

٧٨ ٦٧ ٦٥ ٥٨ ٥٧ ٤١ ٣٧ ٣٠ ٣٠ ٢٨ ٢٦ ١٥  
ب أ أ ب ب أ أ أ أ ب أ أ

١٩٨١٤١١٣٧١٣٥١٣٢١٢٢١٢١١٠٣٩١ ٨٠  
أ ب ب أ أ ب ب ب ب ب ب

الفرض الصفري : المعالجة ليس لها تأثير في نمو البراعم . أى أن البيانات مأخوذة من مجتمع واحد وبالتالي تتابع الرموز ا ، ب عشوائيا .

الفرض الآخر : المعالجة تعيق نمو البراعم ، وإذن الاختبار ذو جانب واحد .

إذا كان الفرض الصفري صحيحا فإن جميع المتابعات من الرمز ا ، ب التي يمكن أن تسفر عنها التجربة تكون متساوية الاحتمال ويتوزع هذان الرمان عشوائيا . أما إذا كانت المجموعتان هما عيتتان من مجتمعين مختلفين نتيجة لتأثير المعالجة بالهورمون فإن عناصر كل من المجموعتين تميل إلى التجمع معا ويكون عدد التلاحقات قليلا . وعلى ذلك يمكن استخدام اختبار التلاحقات بالأسلوب السابق بيانه في الأمثلة الثلاثة السابقة .

عدد التلاحقات س = ٩ ، ن<sub>١</sub> = ١١ ، ن<sub>٢</sub> = ١١

من (١) : 
$$١٢ = ١ + \frac{١١ \times ١١ \times ٢}{٢٢} =$$
 ملر



(٣) اختبار عند مستوى الدلالة ٠,٠٥ ما إذا كانت قيم العينة الآتية مرتبة ترتيباً عشوائياً :

٥٧	٢٣	٦٣	٨٤	١٦	٥٥	١٢	١٦	٣	٩٧
٣١	١٦	٩٩	٥٦	٧٠	٣٧	٥٢	٢٣	٢٦	١٨

#### (١٤ - ٢) اختبار الإشارة : SIGN TEST

حين نستخدم اختبار ت لاختبار فرض عن الوسط الحسابي لمجتمع ( البند ٦ - ٦ - ١) وعند اختبار فرض تساوى متوسطى مجتمعين ( البند ٦ - ٦ - ٣) نشترط أن تكون المجتمعات معتدلة . أما إذا كان هذا الشرط غير متحقق ولا يمكن الدفاع عنه فلا مفر من اللجوء إلى الاختبارات غير البارامترية . ولعل أسهل وأسرع اختبار لذلك هو الاختبار المعروف باختبار الإشارة ، وهو يبنى على توزيع ذى الحدين .

#### (١٤ - ٢ - ١) اختبار فرض عن متوسط مجتمع :

نفرض أننا حصلنا على عينة عشوائية من مجتمع معين ونريد أن نختبر ما إذا كان لهذا المجتمع وسط حقيقي معين  $\mu = a$  . يبدأ اختبار الإشارة بوضع الإشارة + بدلا من كل قيمة في العينة تزيد عن  $a$  ووضع الإشارة - بدلا من كل قيمة تقل عن  $a$  وإهمال القيم التي تساوى  $a$  . إذا كان الفرض الصفرى  $\mu = a$  صحيحاً وكان المجتمع متجانساً نتوقع أن يكون عدد الإشارات الموجبة مساوياً على وجه التقريب لعدد الإشارات السالبة . أما إذا بدا أن أحدهما أكبر مما ينبغي فإننا نرفض ذلك الفرض الصفرى .

ليكن  $n_+$  ،  $n_-$  رمزين لعددى الإشارات الموجبة والسالبة على الترتيب . إذا كان الفرض صحيحاً فإن احتمال أن تزيد أى قيمة مشاهدة عن العدد  $a$  يساوى

احتمال أن تقل عن ١ وعلى ذلك فإن كلا الاحتمالين يساوي  $\frac{1}{2}$  ولذلك فإن اختبار الإشارة يعرف متغيراً عشوائياً  $\mu$  يعبر عن عدد الإشارات الموجبة (أو السالبة) في  $n$  من العناصر. وإذا كانت القياسات مستقلة يكون لهذا المتغير توزيع ذي الحدين دليلاً  $n$ ،  $\frac{1}{2}$  حيث  $n$  هو حجم العينة بعد استبعاد القيم التي أهملت. ولوضع قاعدة روتينية لهذا الاختبار نرمز إلى قيم هذا المتغير بالرمز  $S$  حيث  $S$  أصغر العددين  $n$ ،  $n - S$  أي أن :

$$S = \min(n, n - S) \text{ أصغر } (n, n - S)$$

ثم نحسب احتمال أن يأخذ المتغير  $\mu$  قيمة تساوي أو تقل عن القيمة المشاهدة  $S$  أي نحسب الاحتمال :

$$L(S \geq S) \quad (٧)$$

على أساس صحة الفرض الصفري أن  $\mu = \frac{1}{2}$  (وبالتالي  $\mu = 1$ ). وإذا كانت  $\alpha$  هي مستوى الدلالة الذي اخترناه فإننا نرفض الفرض الصفري عند هذا المستوى في حالة الاختبار ذي الجانب الواحد إذا كان :

$$L(S \geq S) > \alpha \quad (٨)$$

ونقبله إذا زاد هذا الاحتمال عن  $\alpha$  أو كان مساوياً لها. وفي حالة الاختبار ذي الجانبين نرفض الفرض الصفري إذا كان :

$$L(S \geq S) > \alpha \quad (٩)$$

وحين يكون حجم العينة صغيراً ( $n \geq 15$ ) نوجد الاحتمال (٧) مباشرة من أحد جداول احتمالات توزيع ذي الحدين كالجدول (٣) بملحق هذا الكتاب، أما إذا كانت  $n$  أكبر من ١٥ ولا يتسع لها هذا الجدول فإننا نستخدم تقريب التوزيع المعتدل لتوزيع ذي الحدين الذي مر بنا بالبند (٤ - ٦) إذا توفرت شروطه، ونستخدم نفس مناطق الرفض كما في البند (١٤ - ١) الأخير.

مثال (١٤ - ٥) :

الأعداد الآتية هي ١٥ قياساً لمعدلات الأوكتين في نوع من الجاسولين :

٩٧ ٩٤,٧ ٩٦,٨ ٩٥,٥ ٩٦,٣ ٩٩,٨ ٩٦,٩ ٩٤,٨ ٩٢,٧  
٨٦,٦ ٩٥,٦ ٩٧,١ ٩٧,٧ ٩٨,٠ ٩٤,٤

اختبر عند مستوى الدلالة ٠,٠١ الفرض الصفري أن متوسط معدل الأوكتين هو  $\mu = 98$  ضد الفرض الآخر أن  $\mu > 98$ .

الحل :

باستخدام قاعدة الإشارات سابقة الذكر نحصل على الإشارات الآتية وعددها ١٤ إشارة بعد إهمال العدد الذي يساوى ٩٨ .

-----+-----+-----

إذن  $u_+ = 2$  ،  $u_- = 12$  من ١٤ إشارة .  
نأخذ  $S = 2$  ( أصغر العددين ٢ ، ١٢ ) .

نختبر الفرض أن توزيع عدد الإشارات الموجبة هو توزيع ذى الحدين دليلاً  
١٤ ، ٠,٥ . إذا كان هذا الفرض صحيحاً نجد من الجدول ( ٣ ) عند  $n = 14$  ،  
 $H = \frac{1}{4}$  أن :

$$L (S \geq 2) = 0,001 + 0,006 = 0,007$$

وهذا الاحتمال أصغر من مستوى الدلالة ٠,٠١ ولهذا نرفض الفرض الصفري ونستنتج أن  
متوسط معدل الأوكتين يقل عن ٩٨ .

ملاحظة (١) :

استخدام تقريب التوزيع المعتدل :

في هذا المثال نظراً لأن  $n = 2 = (n - 1) = 1$  وهذا العدد أكبر من خمسة ، وحسب الإحصاءة (٦) بالبند (٦ - ٤) يمكن تقريب توزيع ذى الحدين الذى لدينا بالتوزيع المعتدل عن طريق الإحصاءة :

$$\mu = \frac{\mu - (\frac{1}{2} \pm \sigma)}{\sigma}$$

حيث  $\mu = n = 2$

$$\sigma = \sqrt{n} = \sqrt{2} = 1.414$$

وعلى أساس صحة الفرض نجد أن :

$$2.406 = \frac{2 - (\frac{1}{2} \pm \sigma)}{1.414}$$

وهذا العدد يقل عن - 2.33 وإذن نرفض الفرض الصفري عند المستوى 0.01

ملاحظة (٢) :

نؤاقع أن الذى يختبره اختبار الإشارة هو الوسيط ولكن نظراً لأننا افترضنا أن المجتمع متماثل فإن الوسيط يكون هو نفسه الوسط الحسابى ، أما إذا لم يكن المجتمع متماثلاً فإن اختبار الإشارة يكون اختياراً عن الوسيط وليس عن الوسط الحسابى .

(١٤ - ٢ - ٢) مقارنة متوسطة مجموعتين غير معتلين :

نفرض أن لدينا مجموعة من أزواج المشاهدات (س١ ، ص١) من مجموعتين

متصلين غير معتدلين ونريد اختبار ما إذا كان لهذين المجتمعين وسطان حسايان متساويان  $(\mu_1 = \mu_2)$ . بنفس أسلوب البند السابق ، يبدأ اختبار الإشارة بوضع الإشارة + أو الإشارة - لكل فرق  $F_i = \text{سر} - \text{صر}$  بحسب كون هذا الفرق موجباً أو سالباً ، وإهمال الحالات التي ينعدم فيها هذا الفرق .

إذا كان الفرض الصفري  $\mu_1 = \mu_2$  صحيحاً وكان كل من المجتمعين متماثلاً ينبغي أن يكون مجموع الإشارات الموجبة في العينة مساوياً بالتقريب لمجموع الإشارات السالبة ويكون احتمال أن يكون الفرق  $F$  موجباً يساوى احتمال أن يكون هذا الفرق سالباً ويكون كلا الاحتمالين مساوياً للعدد  $\frac{1}{2}$  ولذلك فإن الاختبار يعرف متغيراً عشوائياً له توزيع ذى الحدين دليله  $H = \frac{1}{2}$  بشرط صحة الفرض الصفري . ويسمى الاختبار بنفس الأسلوب المذكور في البند السابق .

مثال (١٤ - ٦) :

في دراسة لمعرفة تأثير نظام جديد في المرور جمعت البيانات الآتية عن عدد الحوادث التي وقعت في ١٢ تقاطعاً من التقاطعات الخطرة ( على فرض أنها عينة عشوائية ) خلال ٣ شهور قبل النظام الجديد و ٣ شهور بعده :

قبل النظام الجديد : ٣ ٥ ٢ ٣ ٣ ٠ ٤ ١ ٥ ٤ ١

بعد النظام الجديد : ١ ٢ ٠ ٢ ٢ ١ ٢ ٣ ٣ ٣ ٠

ابحث عند مستوى الدلالة ٠,٠٥ ما إذا كان النظام الجديد له أثر في تقليل الحوادث عند تلك التقاطعات .

الحل :

باستخدام قاعدة الإشارات للفروق نحصل على الإشارات الآتية وعددها ١٢

++++-+-+++++

$\text{+} = ١٠$  ،  $\text{-} = ٢$  من ١٢ إشارة



الفرض الصفري ف:  $\mu = \mu_0$  متوسط عدد الحوادث واحد في الحالتين

$$\mu_1 < \mu_0$$

نأخذ س = ٢ (أصغر العددين ١٠ ، ٢)

نختبر الفرض أن توزيع عدد الإشارات السالبة هو توزيع ذي الحدين دليلاً  
١٢ ،  $\frac{1}{4}$  - إذا كان هذا الفرض صحيحاً نجد من الجدول (٢) عند  $h = ١٢$  ،  
 $\frac{1}{4} = \frac{1}{4}$  أن :

$$L = (S \geq ٢) = ٠,٠٠٣ + ٠,٠١٦ = ٠,٠١٩$$

وهذا الاحتمال أقل من ٠,٠٥ وإذن نرفض الفرض الصفري عند مستوى الدلالة  
٠,٠٥ لصالح الفرض الآخر ونحكم بأن النظام الجديد قد أثر في تقليل عدد  
الحوادث عند التقاطعات الخطرة .

## تمارين (١٤ - ٢)

(١) في إحدى التجارب العملية نتجت ١٨ قيمة لمعامل الاحتكاك بين الجلد  
وأحد المعادن وكانت هذه القيم كالآتي :

٠,٥٩ ٠,٤٨ ٠,٥٣ ٠,٥٧ ٠,٥٥ ٠,٦١ ٠,٥٦ ٠,٦٠ ٠,٥٥  
٠,٥٨ ٠,٥٢ ٠,٥٥ ٠,٥٦ ٠,٥١ ٠,٥٧ ٠,٥٦ ٠,٥٩ ٠,٥٨

استخدام اختبار الإشارة عند مستوى الدلالة ٠,٠٥ لاختبار الفرض الصفري أن  
متوسط معامل الاحتكاك  $\mu = ٠,٥٥$  ضد الفرض الآخر أن  $\mu \neq ٠,٥٥$

(٢) الآتي هي أعداد الحفريات التي وجدها اثنان من علماء الآثار في بقايا  
مساكن أثرية على سفح جبل في ٣٠ يوماً :

العالم الأول : ١ ٠ ٢ ٣ ١ ٠ ٢ ٢ ٣ ١ ٠ ١ ٤ ١ ٢  
العالم الثاني : ٠ ١ ٠ ٢ ٠ ١ ٠ ١ ٢ ١ ٠ ١ ٢ ١ ١

العالم الأول : ٢ ٤ ٢ ٠ ٢ ٣ ١ ٤ ٢ ٣ ١ ٢ ٥ ٣ ٠ ١  
العالم الثاني : ٠ ١ ٠ ١ ٠ ١ ٢ ٠ ٣ ٢ ٠ ٦ ٢ ٢ ٠

استخدم اختبار الإشارة عند مستوى الدلالة ٠,٠١ لاختبار الفرض الصفري  
أن العالمين على نفس الكفاءة في العثور على الحفريات ضد الفرض الآخر أن العالم  
الأول أفضل .

(٣) أعطى كل من ١٠ مرضي نوعان من المهدئات أ ، ب والجدول الآتي  
يعرض الزيادة في مدة النوم بالساعات . هل الفرق بين نوعي المهدئات ذو دلالة ؟

أ : ٣,٤ ٤,٦ ١,٦ ٥,٥ ٤,٤ ٠,١ - ٠,١ ١,١ ٠,٨ ١,٩  
ب : ٢,٠ ٠,١ ٠,٨ ٣,٧ ٣,٤ ٠,١ - ١,٢ - ٠,٢ - ١,٦ - ٠,٧

(١٤ - ٣) اختبار ويلكوكسن للمقارنات التزاوجية :

#### WILCOXON TEST FOR PAIRED COMPARISONS

إن اختبار الإشارة الذي ورد بالبند (١٤ - ٢ - ٢) يحدد أيًا من المجتمعين  
المأخوذة منهما العينات هو الأكبر في المتوسط ولكنه لا يحدد مقدار الفرق بينهما .  
والاختبار الذي يحدد كلا الاتجاه والمقدار هو ذلك المعروف باختبار ويلكوكسن  
للمقارنات التزاوجية ، وهو فضلًا عن هذا أكثر حساسية من اختبار الإشارة في  
الكشف عن وجود فرق بين متوسطي مجتمعين توزيعهما مجهولان . وتتضح أهمية  
هذا الاختبار في الحالات التي لا تنطبق فيها شروط الاختبار المتقدم في البند  
(٨ - ٧) .

نفرض أننا حصلنا على  $n$  من أزواج القيم ( $x_i$  ،  $y_i$ ) . وليكن  $F_i = x_i - y_i$  - صفر  
هي الفروق بين هذه الأزواج . لاختبار الفرض الصفري  $\mu_x = \mu_y$  عن  
تساوي متوسطي المجتمعين يبدأ اختبار ويلكوكسن بإهمال الفروق المساوية للصفر

ثم ترتيب الفروق الباقية بصرف النظر عن إشاراتها أى بحسب القيم المطلقة لهذه الفروق ، فتعطى الرتبة ١ لأصغر فرق في القيمة المطلقة وتعطى الرتبة ٢ للفرق التالى في الصغر وهكذا . وحين تتساوى القيم المطلقة لاثنتين أو أكثر من هذه الفروق يعطى لكل منهما متوسط الرتب التي كانت ستعطى لو كانت هذه الفروق متميزة .

إذا كان الفرض الصفري  $\mu_0 = \mu_1$  صحيحا نتوقع أن يكون مجموع الرتب المناظرة للفروق الموجبة في العينة مساويا بالتقريب لمجموع الرتب المناظرة للفروق السالبة . ل نرمز إلى هذين المجموعين بالرمزين  $T_1$  ،  $T_2$  على الترتيب ، وليكن

$$S = \text{أصغر } (T_1, T_2)$$

نظراً لأن  $S$  تتغير من عينة إلى أخرى فإننا ننظر إليها على أنها قيمة مشاهدة من متغير عشوائى  $S$  . إن هذا المتغير له توزيع معروف متوسطه وتباينه كالآتي :

$$\mu_S = \frac{n(n+1)}{4} \quad (10)$$

$$\sigma_S^2 = \frac{n(n+1)(n+2)}{24} \quad (11)$$

وعلى ذلك تبني قاعدة الاختبار على إيجاد احتمال الحصول بالصدفة وحدها على قيمة تساوى أو تقل عن قيمة  $S$  المشاهدة أى على إيجاد الاحتمال  $L(S \geq S)$  . ونظراً لصعوبة إيجاد هذا الاحتمال فقد أعدت جداول لإعطاء القيم الحرجة لتوزيع المتغير  $S$  ومنها الجدول (١٤) بملحق هذا الكتاب وهو يعطى القيمة الحرجة عند  $\alpha = 0.05$  ،  $0.01$  ،  $0.001$  ،  $0.0001$  عند كل من مستويات الدلالة  $0.1$  ،  $0.05$  ،  $0.01$  ،  $0.001$  للاختبارات ذات الجانب الواحد وعند كل من مستويات الدلالة  $0.05$  ،  $0.01$  ،  $0.001$  للاختبارات ذات الجانبين ، وأى قيمة مشاهدة  $S$  تقل عن القيمة المذكورة في الجدول تكون ذات دلالة عند المستوى المذكور وتدعو إلى رفض الفرض الصفري وذلك لأن هذا الجدول يعطى قيم  $S$  حيث الاحتمال  $L(S \geq S)$  يكون أصغر من مستوى الدلالة المذكور .

وفي الحالة التي تزيد فيها  $n$  عن ٣٠ ولا يتسع لها الجدول (١٤) نستخدم الإحصاءة .

$$(12) \quad \frac{\bar{m} - \frac{n}{2} (1 + n)}{\sqrt{\frac{n}{24} (1 + n) (1 + n^2)}} = \bar{m}$$

التي يقترب توزيعها من التوزيع المعتدل المعياري .

مثال (١٤ - ٧) :

الأعداد المدونة بالعمودين الثاني والثالث من الجدول الآتي هي متوسطات أعداد المواليد في المرة الواحدة لسلالتين من فيران التجارب كانتا محفوظتين في مستعمرات كبيرة في الولايات المتحدة ، وذلك في الأعوام التسعة من ١٩١٦ إلى ١٩٢٤ :

السنة	السلالة (أ)	السلالة (ب)	ف	الرتب (مع إهمال الإشارة)
١٩١٦	٢,٦٨	٢,٣٦	+ ٠,٣٢	٩
١٩١٧	٢,٦٠	٢,٤١	+ ٠,١٩	٨
١٩١٨	٢,٤٣	٢,٣٩	+ ٠,٠٤	٢
١٩١٩	٢,٩٠	٢,٨٥	+ ٠,٠٥	٣
١٩٢٠	٢,٩٤	٢,٨٢	+ ٠,١٢	٧
١٩٢١	٢,٧٠	٢,٧٣	- ٠,٠٣	١
١٩٢٢	٢,٦٨	٢,٥٨	+ ٠,١٠	٦
١٩٢٣	٢,٩٨	٢,٨٩	+ ٠,٠٩	٥
١٩٢٤	٢,٨٥	٢,٧٨	+ ٠,٠٧	٤

اختبر عند مستوى الدلالة ٠,٠١ ما إذا كان متوسط الخلفة للسلالة أ يزيد عنه في السلالة ب .

## الحل :

نحسب الفروق  $F$  بين كل زوج من المشاهدات ( السلالة ١ - السلالة ب ) مع الاحتفاظ بالإشارة الموجبة أو السالبة كما هو مبين بالعمود الرابع . نرتب الفروق من الأصغر إلى الأكبر بصرف النظر عن الإشارة كما هو مبين بالعمود الأخير .

$$\begin{aligned} \text{لدينا } n &= 9 \text{ ( عدد أزواج القيم في العينة )} \\ + &= 44 \text{ ( مجموع الرتب المناظرة للفروق الموجبة )} \\ - &= 1 \text{ ( مجموع الرتب المناظرة للفروق السالبة )} \\ \text{نأخذ } S &= 1 \text{ ( أصغر العددين ( 44 ، 1 ) )} \end{aligned}$$

من الجدول (١٤) وعند  $n = 9$  ،  $\alpha = 0.01$  ، نجد القيمة الحرجة ٣ وأى قيمة  $S$  تساوى أو تقل عن هذه القيمة تكون ذات دلالة عند المستوى ٠,٠١ . وبما أن  $S = 1 \geq 3$  فإنها تكون ذات دلالة وتدعو إلى رفض الفرض الصفري ونستنتج أن متوسط الحلقة في السلالة (أ) أكبر منه في السلالة (ب) .

## ملاحظة (١) :

في هذا المثال يحق لنا اعتبار أننا بصدد مقارنات تزاوجية وذلك بملاحظة توافر التغيرات في السلالتين معاً خلال السنوات التسع . ففي السنتين ١٩١٧ ، ١٩١٨ وهما سنتا حرب في الولايات المتحدة وفي العالم كله ، أدى النقص في الرعاية وفي الغذاء إلى قلة عدد الذرية في السلالتين ثم تحسن هذا العدد بمجرد تحسن الظروف . كذلك نلاحظ أنه في العام ١٩٢٢ كان هناك هبوط في الذرية في كلا السلالتين ، مما يشير إلى أن التغيرات ترجع إلى أسباب بيئية . ولهذا يكون من المناسب تناول هذه البيانات على أنها مقارنات تزاوجية العامل الثابت فيها هو عامل السلالة أما السنوات فهي عامل التكرارات .

## ملاحظة (٢) :

يمكن استخدام اختبار ويلكوكسن للمقارنات التزاوجية لاختبار الفرض

الصفرى أن الفرق بين متوسطى مجتمعين يأخذ قيمة معينة ، مثلاً  $\mu_1 - \mu_2 = 1$  . ولا يختلف الإجراء المطلوب عن الإجراء السابق إلا في أننا نطرح العدد 1 من كل فرق ف قبل إعطاء الرتب كما في المثال الآتى . .

مثال (١٤ - ٨) :

كان هناك ادعاء بأنه إذا أعطى طالب في المستوى الرابع عينات من امتحانات سابقة فإن درجاته عند التخرج تزيد بمقدار ٥٠ نقطة . اختبر ٢٠ طالباً عشوائياً في هذا المستوى وقسموا إلى ١٠ أزواج بحيث كان الطالبان في كل زوج متساويين تقريباً في معدلهما التراكمى عن السنوات الثلاثة السابقة ، وأعطيت عينات من امتحانات سابقة لواحد فقط من كل زوج ( اختبر بطريقة عشوائية ) قبل الامتحان بأسبوع وسجلت درجات الامتحان النهائى في العمودين الثانى والثالث من الجدول الآتى . اختبر عند مستوى للدلالة ٠,٠٥ الفرض الصفرى أن  $\mu_1 - \mu_2 = ٥٠$  ضد الفرض الآخر أن  $\mu_1 - \mu_2 > ٥٠$

الزوج	بامتحانات سابقة	بغير امتحانات سابقة	ف	ف - ٥٠	الرتب (مع إمل الإشارة)
١	٥٣١	٥٠٩	٢٢	٢٨-	٥
٢	٦٢١	٥٤٠	٨١	٣١	٦
٣	٦٣٣	٦٨٨	٢٥-	٧٥-	٩
٤	٥٧٩	٥٠٢	٧٧	٢٧	٣,٥
٥	٤٥١	٤٢٤	٢٧	٢٣-	٢
٦	٦٦٠	٦٨٣	٢٣-	٧٣-	٨
٧	٥٩١	٥٦٨	٢٣	٢٧-	٣,٥
٨	٧١٩	٧٤٨	٢٩-	٧٩-	١٠
٩	٥٤٣	٥٣٠	١٣	٣٧-	٧
١٠	٥٧٥	٥٢٤	٥١	١	١

لدينا  $n = 10$  (حجم العينة) .

$$s^2 = 1 + 3,5 + 6 = 10,5$$

$$s^2 = 7 + 10 + 3,5 + 8 + 2 + 9 + 5 = 44,5$$

نأخذ  $s = 10,5$  (أصغر العددين  $10,5$  ،  $44,5$ )

من الجدول (١٤) وعند  $n = 10$  ،  $\alpha = 0,05$  نجد القيمة الحرجة ١١.

بما أن  $s = 10,5 \geq 11$  نرفض الفرض الصفري ونستنتج أنه ( في المتوسط ) إعطاء الامتحانات السابقة لا يجعل درجة التخرج تزيد بمقدار يصل إلى ٥٠ نقطة .

### تمارين (١٤ - ٣)

(١) بالجدول الآتي الأوزان بالكيلو جرامات لخمس أشخاص قبل أن يمتنعوا عن التدخين وبعد ٥ أسابيع من امتناعهم عنه :

(٥)	(٤)	(٣)	(٢)	(١)	
٥٤	٥٢	٦٩	٨٠	٦٦	قبل
٥٩	٥٦	٦٨	٨٢	٧١	بعد

استخدم اختبار ويلكوكسن للمقارنات التزاوجية عند مستوى الدلالة ٠,٠٥ ، لاختبار أن الامتناع عن التدخين ليس له تأثير في زيادة الوزن ضد الفرض الآخر أنه يزيد الوزن .

### TEST FOR TREND (١٤ - ٤) اختبار الاتجاه :

ليكن  $y$  متغيراً عشوائياً ،  $x$  متغيراً رياضياً ( كما في الفصل التاسع عن الانحدار الخطي البسيط ) . كثيراً ما نتساءل : إذا ازدادت قيم  $y$  فهل تزداد معها قيم  $x$  ( اتجاه موجب ) أم تنقص قيم  $x$  ( اتجاه سالب ) ؟ أي أننا نريد اختبار الفرض الصفري أنه لا يوجد اتجاه ، ضد الفرض الآخر بوجود اتجاه موجب أو سالب أو بوجود اتجاه بصفة عامة .

إذا توفرت الافتراضات المذكورة في الفصل التاسع ( خطية العلاقة بين  $s$  ،  $s$  واعتدالية توزيع المتغير العشوائي  $s$  ) فإننا نتبع الأسلوب المين بذلك الفصل ، أما إذا لم تكن متوفرة فيمكننا أن نستخدم اختباراً بسيطاً يعرف باختبار الاتجاه .

وعلى فرض أن لدينا  $h$  من أزواج القيم  $(s, s)$  ناتجة من عينة عشوائية فإن هذا الاختبار يبدأ بترتيب قيم  $s$  ترتيباً تصاعدياً ثم ملاحظة قيم  $s$  المناظرة . في حالة وجود اتجاه موجب ينبغي أن تزداد قيم  $s$  مع ازدياد  $s$  . أما إذا لم يوجد اتجاه فإن قيم  $s$  تسلك سلوكاً عشوائياً دون أى رتابة أى دون أن يكون هناك نسق معين . والمؤشر لذلك هو عدد الاستبدالات  $transpositions$  في المتغير  $s$  أى عدد المرات التي لا تكون فيها قيم  $s$  في مكانها الطبيعي من التزايد أى حين تسبق بعض هذه القيم قيمة أصغر منها . ولذلك فإن الاختبار يعرف متغيراً عشوائياً  $s$  يعبر عن عدد الاستبدالات ثم يحسب الاحتمال :

$$(13) \quad L(s \geq s) \quad (13)$$

على أساس صحة الفرض الصفري بعلم وجود اتجاه وحيث  $s$  هي عدد الاستبدالات المشاهدة في العينة . فإذا كان هذا الاحتمال أقل من مستوى الدلالة الذي نختاره فإننا نرفض الفرض الصفري عند هذا المستوى لصالح الفرض الآخر . وقد أعدت جداول للاحتالات المتجمعة المبينة بالصيغة (13) ومنها الجدول (15) ملحق هذا الكتاب الذي يعطى هذه الاحتمالات لقيم  $h = 3, 4, \dots, 20$  بالنسبة للمتغيرات المتصلة .

مثال (١٤ - ٩) :

في بحث لمعرفة تأثير تعرض بذور الشعير لنوع من الأشعة على المحصول الناتج وجدت البيانات الآتية مع ملاحظة أن  $s$  وضعت مرتبة ترتيباً تصاعدياً وأن  $s$  مقاسة بالجرامات في الجوال .



مقدار الأشعة (س) : ٠ ٢ أ ٤ أ ٤٠ أ ٤٠٠ أ  
مقدار المحصول (هـ) : ٢٩,٢ ٣٠,٢ ٢٨,٢ ٢٩,٧ ٣٠,٤

اختبر عند مستوى الدلالة ٠,٠٥ ما إذا كان مقدار المحصول يزداد بازدياد مقدار الأشعة .

الحل :

الفرض الصفري ف<sub>٠</sub> : تعريض البنور للأشعة لا يؤثر في المحصول .

الفرض الآخر ف<sub>١</sub> : يوجد اتجاه موجب (ص تزداد بازدياد س) .

نحسب عدد الاستبدالات كآلاتي :

العدد ٢٩,٢ يسبق العدد ٢٨,٢ ( استبدال واحد ) .

العدد ٣٠,٢ يسبق العددين ٢٨,٢ ، ٢٩,٧ ( استبدالين اثنين ) .

إذن عدد الاستبدالات س = ٣ في ه = ٥ من القيم الصادية .

من الجدول (١٥) عند ه = ٥ ، س = ٣ وعلى أساس صحة الفرض الصفري

نجد أن : ل ( س ≥ ٣ ) = ٠,٢٤٢

وهذا الاحتمال أكبر من مستوى الدلالة ٠,٠٥ وعلى ذلك لا نستطيع رفض الفرض

الصفري ونستنتج أنه في حدود هذه التجربة ليس للأشعة تأثير جوهري على

محصول الشعير .

ملاحظة :

إذا اشتملت العينة على ٢ من القيم المتساوية للمتغير ص نحسب عدد الاستبدالات

كما سبق ثم نضيف إليه العدد  $\frac{٢-١}{٤}$  ( ٢ - ١ ) وهو متوسط الاستبدالات في حالة

تبادل ٢ من الأشياء . فمثلا إذا وجدت قيمتان متساويتان أي ٢ = ٢ نضيف

العدد  $\frac{٢-١}{٤} = ١ \times ٢ \times \frac{٢-١}{٤}$  وإذا كانت ٢ = ٣ نضيف العدد  $\frac{٣-١}{٤} = ٢ \times ٣ \times \frac{٣-١}{٤}$

لكل ثلاثة قيم متساوية وهكذا .

## تمارين (١٤ - ٤)

استخدم اختبار الاتجاه لكل من العينات الآتية :

(١)	ص :	١٥	٢٠	٣٠	٤٠	٥٠	٦٠	٧٠
	ص :	٦,٥	٥,٦	٥,٤	٦,٠	٤,٦	١,٤	٠,١

حيث ص هي محتوى الأكسجين ( ملليجرام / لتر ) في أحد البحيرات عند العمق س بالمأتر .

(٢)	ص :	١٠	١٥	٣٠	٤٠	٥٠	٥٥	٨٠	١٠٠
	ص :	٥٠	٤٦	٤٣	٤٣	٣٦	٣٩	٣٧	٣٣

حيث ص هي محتوى الدسلفيد في ميثيل الصوف ،  
ص نسبة تركيز محتوى الماء .

(٣)	ص :	١٢١	١٢٠	٩٥	١٢٣	١٤٠	١١٢	٩٢	١٠٠	١٠٢	٩١
	ص :	٥٢١	٤٦٥	٣٥٢	٤٥٥	٤٩٠	٣٨٨	٣٠١	٣٩٥	٣٧٥	٤١٨

حيث ص تعبر عن ضغط الدم بوحدات معينة ، ص وزن القلب بالجرامات  
لعينة من ١٠ مرضي ذكور ماتوا بأحد أمراض المخ .

## (١٤ - ٥) اختبار كروسكال - واليس KRUSKAL-WALLIS TEST

يستخدم هذا الاختبار كبديل لطريقة تحليل التباين للتجارب ذوات العامل الواحد حين لا تتوفر شروطها .

وتتلخص المشكلة التي نتاولها هنا فيما يلي :

لدينا ك من العينات العشوائية أحجامها  $n_1$  ،  $n_2$  ، ... ،  $n_k$  مأخوذة

من ك من المجتمعات ، ونرغب في اختبار الفرض الصفري ف أن لهذه المجتمعات متوسطات متساوية . إذا كان هذا الفرض صحيحا يمكن النظر إلى هذه العينات على أنها عينة عشوائية واحدة حجمها  $n$  ( هو مجموع أحجام العينات ) مأخوذة من مجتمع مشترك . نرتب قيم هذه العينة من الأصغر إلى الأكبر - من ١ إلى  $n$  - ونستعيض عن كل قيمة مشاهدة بالترتيب المناظر لها . نجمع تراتيب وحدات كل عينة على حدة ، ولتكن  $T_1, T_2, \dots, T_r$  ، هي مجاميع هذه التراتيب . إذا كان الفرض الصفري ف صحيحا فإن هذه المجاميع تكون ذات قيم متقاربة ، أما إذا لم يكن ف صحيحا فإن التراتيب الكبرى تميل إلى أن تقع في العينات المأخوذة من المجتمعات التي لها أكبر المتوسطات . وعلى هذا نكون في حاجة إلى إحصاءة اختبار تكشف عن وجود أو عدم وجود واحد أو أكثر من المجاميع الترتيبية  $T_i$  تكون كبيرة بدرجة لا نتوقع حدوثها عن طريق الصدفة .

وقد وجد أن أحد الإحصاءات التي تصلح لذلك هي تلك التي قدمها كروسكال وواليس وهي تأخذ الصيغة الآتية :

$$H = \frac{12}{n(n+1)} \sum_{i=1}^r \frac{T_i^2}{n_i} - \frac{3}{2} \quad (14)$$

حيث  $n$  مجموع حجوم العينات ،  $T_i$  مجموع تراتيب وحدات العينة  $i$  ،  $n_i$  = ١ ، ٢ ، ... ،  $r$  . ولهذا الإحصاءة توزيع قريب من توزيع  $\chi^2$  بدرجات حرية  $r - 1$  وبالتالي يمكن استخدام جدول  $\chi^2$  لاختبار الفرض الصفري عن تساوي متوسطات المجتمعات ، فرفض ف . إذا كانت القيمة المشاهدة للإحصاءة (١٤) أكبر من القيمة الحرجة في توزيع  $\chi^2$  عند مستوى الدلالة الذي نختاره .

مثال (١٤ - ١٠) :

أراد أحد رجال التربية الرياضية اختبار أفضلية ثلاثة طرق جديدة في تعليم لعبة كرة السلة فأخذ عينة عشوائية من عشرين من المبتدئين في هذه اللعبة وقسمها عشوائيا إلى أربع مجموعات بكل منها خمسة لاعبين . دربت إحدى المجموعات بالطريقة المعتادة بينما دربت كل من المجموعات الثلاث الأخرى بإحدى الطرق الجديدة . وبعد فترة التدريب قيست مهارات اللاعبين وسجلت درجاتهم في الجدول الآتي الذي سجلت فيه أيضا الترتيب المناظرة للدرجات ( وهى تلك الموضوعية بين الأقواس ) . المطلوب اختبار ما إذا كانت هناك فروق ذات دلالة بين مهارات اللاعبين الناتجة عن الطرق الأربع .

الطريقة المعتادة	الطرق الجديدة للتدريب		
د	ح	ب	ا
( ٩ ) ٦٥,٠	( ٢٠ ) ٩٠,٢	( ١٨ ) ٨٥,٠	( ٨ ) ٦٣,٠
( ١ ) ٤٥,٢	( ١١ ) ٧٠,٧	( ١٦ ) ٨٠,١	( ٢ ) ٤٧,٠
( ٣ ) ٥٠,٩	( ١٩ ) ٨٦,٠	( ١٥ ) ٧٩,٠	( ٤ ) ٥١,٠
( ١٤ ) ٧٥,٠	( ٧ ) ٦٢,٣	( ١٠ ) ٦٧,٠	( ١٣ ) ٧٤,٠
( ٥ ) ٥٨,٨	( ١٢ ) ٧٢,٣	( ١٧ ) ٨٢,٣	( ٦ ) ٦٠,٠
( ٣٢ )	( ٦٩ )	( ٧٦ )	ت : ( ٣٣ )

الحل :

الفرض الصفري ف : الطرق الأربع تؤدي في المتوسط إلى نفس الدرجة من المهارات .

من الاحصاءة (١٤) ومع ملاحظة أن  $u = ٢٠$  ،  $k = ٤$  نجد أن

$$ص = \frac{١٢}{٢١ \times ٢٠} \left( \frac{٣٢ + ٦٩ + ٧٦ + ٣٣}{٥} \right) - ٢١ \times ٣ =$$

$$٦٣ - ٢٥٣٠ \times ٠,٠٢٨٥٧ =$$

$$٩,٢٨٢١ = \text{بدرجات حرية } k - ١ = ٣$$

ولكن  $\chi^2_{0,٠٥(٣)}$  ،  $٧,٨١٥$  ، إذن نرفض  $F$  عند مستوى الدلالة  $٠,٠٥$  ونحكم بأن المهارات تختلف باختلاف طرق التدريب . ويدل أن الطريقة ب هي الأفضل .

## FRIEDMAN TEST

(١٤ - ٦) اختبار فريدمان

يستخدم هذا الاختبار كبديل لطريقة تحليل التباين للتجارب التي تصمم على هيئة قطاعات كاملة التعشية أو للتجارب ذوات العاملين غير المتفاعلين ، حين لا تتوفر شروطها .

نفرض أن لدينا  $k$  من المعالجات ونرغب في اختبار ما إذا كان لهذه المعالجات تأثيرات مختلفة على وحدات متغير ما ، ونفرض أنه عند تطبيق هذه المعالجات على وحدات التجريب يدخل عامل خارجي له  $h$  من المستويات قد يؤثر في النتائج التي نحصل عليها وينبغي إذن استبعاد أثر هذا العامل . لتحقيق هذا الفرض نقوم بتعشية هذا العامل بحيث يكون لكل مستوى من مستويات العامل الذي ندرسه نفس الفرصة للتعرض له . فنسحب  $k$  من الوحدات من كل من الـ  $h$  مجتمعات التي تمثل مستويات هذا العامل لنحصل على  $h$  من القطاعات حجم كل منها  $k$  ، ثم نقوم بتطبيق الـ  $k$  معالجات عشوائياً على وحدات كل قطاع ثم نسجل المشاهدات في جدول ذي  $k$  من الأعمدة تمثل المعالجات ،  $h$  من الصفوف تمثل القطاعات .

تبدأ طريقة فريدمان بترتيب المشاهدات في كل قطاع ( صف ) على حدة من

١ إلى ك بحيث يعطى الترتيب ١ للمشاهدة الأصغر ويعطى الترتيب ك للمشاهدة الأكبر . إذا كان الفرض الصفري فـ صحيحا أى كانت المعالجات لها تأثيرات واحدة على المتغير ، ومع ملاحظة أن المشاهدات فى كل قطاع يمكن اعتبارها عينة عشوائية حجمها ك من مجتمع واحد ، فإن الترتيب العالية تتوزع بين مختلف الأعمدة فى مختلف الصفوف . أما إذا كان فـ غير صحيح فإن الترتيب العالية تميل إلى التجمع فى العمود الذى يمثل المعالجة ذات المتوسط الأكبر .

نجمع الترتيب فى كل عمود ولنرمز بالرمز  $T_i$  لمجموع ترتيب العمود  $i$  (١٩)  $= 1, 2, \dots, K$  . نريد احصاءة اختيار تكشف عن وجود أو عدم وجود واحد أو أكثر من المجاميع الترتيبية تكون كبيرة بدرجة لا تحدث بالصدفة . وقد وجد أن إحدى الاحصاءات الصالحة لهذا الغرض هى تلك التى قدمها فريدمان وهى تأخذ الصيغة الآتية :

$$S = \frac{12}{K(K+1)} \sum_{i=1}^K T_i^2 - \frac{K(K+1)}{2} \quad (١٥)$$

حيث ك عدد الأعمدة ، ه عدد الصفوف ،  $T_i$  مجموع الترتيب فى العمود  $i$  . وتوزيع هذه الاحصاءة قريب من توزيع  $\chi^2$  بدرجات حرية ك - ١ وبالتالي نستطيع استخدام جدول  $\chi^2$  فنرفض الفرض الصفري فـ إذا كانت القيمة المشاهدة لهذه الاحصاءة أكبر من القيمة الحرجة لتوزيع  $\chi^2$  عند مستوى الدلالة الذى نختاره .

مثال (١٤ - ١١) :

أراد فريق أبحاث المستهلك فى أحد مصانع فرامل الدرجات مقارنة ثلاثة أنواع من الفرامل التى يتجهها . ولما كان نوع الدراجة التى تستخدم فى التجربة قد يؤثر

في نتائج الدراسة فقد رؤى استبعاد أثر هذا العامل باستخدام تصميم القطاعات كاملة التعشية . اختير ٦ أنواع من الدرجات لتكوين ٦ قطاعات بكل منها ٣ درجات ووزعت الأنواع الثلاثة من الفرامل عشوائيا على كل قطاع . المتغير الذى تقارن به الفرامل هو الزمن بالأسابيع الذى يمضى حتى يتطلب الأمر اصلاحا أساسيا فيها . وقد سجلت هذه الأزمان في الجدول الآتى ، كما رتبت هذه الأزمان في كل قطاع على حدة ووضعت الترتيب المناظرة بين الأقواس في نفس الجدول .

نوع الدراجة القطاعات	أ	ب	ح
(١)	٥,٢ ( ٢ )	٧,٣ ( ٣ )	٣,٠ ( ١ )
(٢)	٦,٨ ( ١ )	٨,٩ ( ٣ )	٧,٥ ( ٢ )
(٣)	٦,٣ (٢,٥)	٦,٣ (٢,٥)	٦,٠ ( ١ )
(٤)	١٣,٠ (١,٥)	١٤,٨ ( ٣ )	١٣,٠ (١,٥)
(٥)	١٢,٨ (٢,٥)	١٢,٨ (٢,٥)	١١,٠ ( ١ )
(٦)	١٥,٠ ( ٢ )	١٥,٢ ( ٣ )	١٤,٥ ( ١ )
ت	(١١,٥)	(١٧)	(٧,٥)

الحل :

الفرض الصفري ف : الأنواع الثلاثة من الفرامل ذات عمر واحد .

من (١٥) ، لدينا  $k = ٣$  ،  $h = ٦$  .

$$\text{ص.} = \frac{12}{4 \times 3 \times 2} \times \left[ \frac{4 \times 6}{2} - 3 \right]$$

$$= \frac{1}{4} [ (12 - 7,5) + (12 - 17) + (12 - 11,5) ]$$

$$= 7,58 = 45,5 \times \frac{1}{4}$$

ولكن  $\chi^2_{[3], 0.05} = 0,991$  ، إذن نرفض الفرض الصفري عند مستوى الدلالة 0,05 ونحكم بأن عمر الفراميل يختلف باختلاف نوعها ، ويدلو أن النوع ب هو الأفضل .

#### تمارين (١٤ - ٥)

١ - أخذت عينات عشوائية من الحجم خمسة من ثلاثة أنهار كبيرة فتحت عنها البيانات الآتية عن مستوى التلوث . استخدم طريقة كروسكال - واليس لبحث ما إذا كانت مستويات التلوث واحدة في الأنهار الثلاثة . استخدم مستوى الدلالة 0,10

النهر الأول	النهر الثاني	النهر الثالث
٢,٧	٢,٩	٠,٦
١,٤	٢,٤	١,٢
٢,٠	٣,٧	١,٥
١,٢	١,٦	١,٧
٢,١	٢,٤	٢,١



٢ - لدراسة تأثير لون الإعلان في جذب الزبائن قامت إحدى الشركات بتصميم خمسة عروض اعلانية متشابهة في كل شيء ما عدا اللون المستخدم ، ثم اختارت عشرة أشخاص عشوائيا وطلب إلى كل منهم ترتيب هذه العروض بحسب مدى جاذبيتها للعين بحيث يعطى الترتيب ١ لأكثر العروض جاذبية والترتيب ٥ لأقلها جاذبية ، فجاءت النتائج كما في الجدول الآتي . هل هناك دليل ( عند المستوى ١٠، ٠) على أن اللون تأثير في جذب الزبائن ؟

(١) بدون ألوان	اللون السائد				الأشخاص (القطاعات)
	(٢) الأحمر	(٣) الأصفر	(٤) الأزرق	(٥) خليط	
١	٢	٤	٣	٥	(١)
١	٢	٥	٣	٤	(٢)
٣	٤	٥	١	٢	(٣)
١	٢	٣	٥	٤	(٤)
٤	٥	١	٣	٢	(٥)
١	٢	٣	٤	٥	(٦)
٢	١	٥	٤	٣	(٧)
١	٢	٤	٥	٣	(٨)
٣	٥	١	٢	٤	(٩)
١	٢	٥	٤	٣	(١٠)



## الفصل الخامس عشر

### اختيار العينات وتحليلها

#### SELECTION AND ANALYSIS OF SAMPLES

يقدم هذا الفصل بعض طرق اختيار العينات وكيفية تحليل ما ينجم عنها من بيانات لتقدير خواص المجتمعات التي أخذت منها .

ولقد ذكرنا في مستهل هذا الكتاب أن المجتمع الإحصائي هو مجموعة من الأشياء أو الأحداث التي تكون موضع اهتمامنا في وقت ما من حيث متغير ما أو عدة متغيرات ، وأشرنا إلى أن دراسة مجتمع ما تقتضى أن يكون هذا المجتمع معرّفا تعريفًا واضحًا خاصة فيما يتعلق بالمتغيرات التي ندرسها وطريقة قياسها وفي تحديد الوحدات التي يتكون منها المجتمع . ونظرًا لأنه من الصعب بل قد يكون من المستحيل دراسة المجتمع بكامله فإن هذه الدراسة تقوم في أغلب الحالات من خلال عينات تختار بحسب خطط معينة تتفق مع طبيعة المجتمع والهدف من دراسته .

والعينة لا تكون ذات قيمة إلا بالقدر الذي تمكّنا به من إصدار أحكام عن الثوابت الإحصائية للمجتمع الذي أخذت منه ، ومن ثم كانت ضرورة العناية القصوى باختيار العينة التي تمثل المجتمع أفضل تمثيل ممكن وبحيث تتسم بصفات تسمح بتحقيق هذا الغرض .

#### (١٥ - ١) المعاينة العشوائية :

من المتطلبات الرئيسية لعملية الاستدلال الإحصائي أن تكون المعاينة من المجتمع عشوائية بمعنى أن تختار العينة بخطة تضمن عدم وجود تحيز من أى نوع قد يؤثر

في اختيارها . ولا يغرب عن بالنا أننا حين نختار عينة عشوائية ما من حجم ما من مجتمع ما بخطة ما فإنما نكون قد اخترنا واحدة من العينات العديدة التي يمكن أن تختار بنفس الخطة وب نفس الحجم من هذا المجتمع . وإذا كانت  $A$  هي التقدير الذي وجدناه في إحدى العينات لأحد ثوابت المجتمع - كالوسط الحسابي - فإن  $A$  تكون واحدة من القيم العديدة التي توجد في العينات الأخرى والتي يمكن أن نقدر بها نفس الثابت . ولذلك نعتبر أن  $A$  هي إحدى قيم متغير عشوائى يهمن أن نعرف توزيع احتماله لأن هذا التوزيع هو الذي نركز عليه في بناء اختبارات الدلالة وتحديد فترات الثقة وتقدير درجات الثقة فيما نصلره من قرارات عن المجتمع ، مما يدخل في موضوع الاستدلال الاحصائي . ولقد سبق الإشارة إلى ذلك في أكثر من مناسبة .

## PROBABILITY SAMPLING (١٥ - ٢) المعاينة الاحتمالية

المعاينة الاحتمالية مصطلح عام يطلق على خطط المعاينة العشوائية التي تختار فيها العينة بحيث يكون لكل وحدة من وحدات المجتمع احتمال معروف للدخول فيها وبحيث تتفق طريقة الاختيار مع هذه الاحتمالات .

إن معرفة هذه الاحتمالات هي التي تتيح لنا استخدام قواعد ونظريات الاحتمال لاستنباط توزيعات الاحتمال اللازمة لعملية الاستدلال الإحصائي .

أما إذا كانت المعاينة غير عشوائية أو كانت احتمالات بعض أو كل وحدات المجتمع للدخول في العينة لا يمكن تحديده فإن العينة تكون حينئذ غير احتمالية . وفي هذه الحال لا نستطيع استخدام الاختبارات الإحصائية أو القيام بعملية الاستدلال الإحصائي بالطرق التي مرت بنا في الفصول السابقة . ومع ذلك لا يجوز التقليل من أهمية مثل هذه العينات التي يمكن الاستفادة منها بطرق أخرى . وهناك عدة خطط للمعاينة الاحتمالية ، نتناول منها هنا الخطط الأكثر شيوعاً في مختلف الميادين التطبيقية وهي : المعاينة العشوائية البسيطة - معاينة طبقية -

المعاينة متعددة المراحل - المعاينة المنتظمة - المعاينة المساحية . وتوقف الخطوة التي نختارها للدراسة مجتمع ما على طبيعة هذا المجتمع من ناحية وعلى نوع الاستنتاجات التي نريد أن نخرج بها عنه من ناحية أخرى . على أن المعيار الرئيسي الذي يجب أن نضعه نصب أعيننا في هذا الاختيار هو الحصول على أكبر قدر ممكن من الدقة في الحكم على المجتمع بأقل مجهود ممكن وبأقل تكلفة .

وينبغي أن تعد خطة المعاينة بالكامل قبل القيام بالتجربة والتجميع الفعلي للبيانات ونبحث تتضمن قاعدتين : قاعدة لطريقة سحب العينة من المجتمع ، وقاعدة بتقدير ثوابت المجتمع من البيانات التي نحصل عليها من العينة مع تقدير مدى الدقة في هذه التقديرات .

سنفترض هنا تسهيل الدراسة أن المجتمعات متبينة وإن كان أغلب المجتمعات ذات أعداد غير متبينة من الوحدات ، وسنهتم بصفة خاصة بتقدير ثابتين هما (١) الوسط الحسابي للمجتمع ذي متغير كمي ، (٢) سببه المجتمع ذي متغير نوعي ذي حددين . ونضع كل من هذين التقديرين تقدير مجموع الكلي نقيم متغير في المجتمع مع العينة بتقدير مدى ثقة في كل من هذه التقديرات

### (١٥ - ٣) العينة العشوائية البسيطة SIMPLE RANDOM SAMPLE

إن هذا النوع من عينات هو أهم أنواع العينات الاحتمالية وأبسطها وينحدر أسماءه من كثير من حفظ العينات الأخرى ، ولقد سمي " قديم عينة عشوائية بسيطة" (١ - ٢) حيث عرفه بأنه تمت عينة عشوائية واحدة من مجتمع حسب كتاب كل واحد من وحداته حيزاً متساوياً واحداً في العينة ، ويجب أن يكون كل واحد من وحداته مستقلاً عن الوحدات الأخرى التي قد تكون فيه ، ولكن يربط أن حيزه بعينه عشوائية بسيطة لعينات التي من حجم معين به بعض كبر مجموعة من به من وحدات المجتمع نفس الفرصة لتكوين عينة .

نجد هذا النوع من أخذ العينة العشوائية البسيطة

إن تعريف العينة العشوائية البسيطة يتضمن أن يكون اختيار العينة متروكا للصدفة وحدها . ولهذا فإن هذه العينة تكون مناسبة إذا كان المجتمع الذى تسحب منه متجانسا من حيث المتغير الذى نتناوله . وإذا كان المجتمع ذا حدين فإن المعاينة العشوائية البسيطة تكون مناسبة إذا كانت النسبة  $\pi$  - وهى احتمال وقوع أى وحدة من وحدات المجتمع فى أحد قسمى المجتمع - واقعة بين ٢٠٪ و ٨٠٪ .

ولا نحتاج فى هذه المرحلة لأى أسس جديدة فى تناول العينات العشوائية البسيطة فقد كانت هى التى نتناولها طوال دراستنا فى الفصول السابقة . على أنه من المهم أن نتذكر دائما أنه إذا أخذت عينة عشوائية بسيطة من مجتمع متوسطه  $\mu$  وتباينه  $\sigma^2$  فإن توزيع المعاينة للأوساط الحسابية للعينات التى من الحجم  $n$  يكون متوسطه  $\mu = \mu$  وتباينه  $\sigma^2 = \frac{\sigma^2}{n}$  راجع البند (٦ - ٢) - وإذا كان المجتمع معتدلا فإن توزيع المعاينة هذا يكون معتدلا : مع  $(\frac{\sigma}{\mu}, \frac{\sigma}{\mu})$  ، وإذا لم يكن المجتمع معتدلا وكان حجم العينة كبيرا فإن توزيع المعاينة يقترب من هذا التوزيع المعتدل كلما زاد حجم العينة - راجع البند (٦ - ٣) . وهذه الحقيقة تيسر لنا التوصل إلى الاستنتاجات التى تتعلق بمتوسطات المجتمعات ومجاميعها .

### (١٥ - ٣ - ١) تقدير الوسط الحسابى والمجموع :

اعتبر مجتمعا حجمه  $N$  ومتوسطه  $\mu$  وتباينه  $\sigma^2$  . إن المجموع الكلى لوحدات هذا المجتمع هو  $M = N\mu$  . نريد تقدير كل من  $\mu$  ،  $M$  من عينة عشوائية بسيطة  $s_1, s_2, \dots, s_n$  ، حجمها  $n$  مأخوذة من هذا المجتمع . تعلم ما يلى :

(أولا) إذا كان متوسط العينة  $\bar{s}$  حيث  $\bar{s} = \frac{1}{n} \sum_{i=1}^n s_i$

فإن هذا المتوسط هو تقدير غير متحيز للمتوسط  $\mu$  للمجتمع . وبالتالي فإن

$$(١) \quad \bar{s} = \mu$$

هو تقدير غير متحيز للمجموع  $M$  للمجتمع .

(ثانيا ) إذا كان تباين العينة  $\sigma^2$  حيث  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (s_i - \bar{s})^2$

فإن  $\sigma^2 = (1 - \frac{n}{N})$  يكون تقديرا غير متحيز للتباين  $\sigma^2$  للمجتمع ، ومن هذا نستطيع اثبات أن  $\sigma^2 = (1 - \frac{n}{N})$  ،  $\sigma^2 = (1 - \frac{n}{N})$  هما على الترتيب تقديران غير متحيزين للتباين لتوزيع المعاينة للمتوسطات وتوزيع المعاينة للمجاميع للعينات ذوات الحجم  $n$  .

ويقدر الخطأ المعياري للمتوسطات بالمقدار  $\sigma = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{N-1}} \sqrt{\frac{N}{N-1}}$  (٢)

كما يقدر الخطأ المعياري للمجاميع بالمقدار  $\sigma = \sigma \sqrt{\frac{N}{N-1}}$  (٣)

وهذان التقديران متحيزان تحيزا قليلا ولكننا نتجاوز عن ذلك في معظم التطبيقات .

يلاحظ أنه إذا كان هناك تقدير غير متحيز لتباين توزيع ما فإن جذره التربيعي ليس من الضروري أن يكون تقديرا غير متحيز للانحراف المعياري للتوزيع .

ويعرف العامل  $\sqrt{1 - \frac{n}{N}}$  أو مربعه بأنه عامل التصحيح للمجموعات المنتهية Finite population correction factor ويمكن إهماله إذا كانت النسبة  $\frac{n}{N}$  ( وهى نسبة حجم العينة إلى حجم المجتمع ) تقل عن حوالى ١٠٪ لأن عامل التصحيح يكون في هذه الحال قريبا من الواحد الصحيح . ويلاحظ من (٢) و(٣) أن كلا من الخطأين المعياريين  $\sigma$  ،  $\sigma$  يساوى صفرا إذا كان  $n = N$  وهذا ما يجب أن يكون لأننا في هذه الحال نكون قد استخدمنا جميع وحدات المجتمع ولا مجال للحديث عن توزيعات المعاينة أو الأخطاء المعيارية .

مثال (١٥ - ١) :

جمعت توقعات على التماس ما في ٦٧٦ بطاقة ، وكانت كل بطاقة قد أعدت لتكفى ٤٢ توقعا، غير أن بعض البطاقات اشتملت على عدد من التوقعات يقل

عن ٤٢ . أخذت عينة عشوائية بسيطة من ٥٠ بطاقة ( أى بواقع حوالى ٧,٤ ٪ ) وحسب عدد التوقعات بكل منها ووضعت النتيجة فى التوزيع التكرارى المبين بالجدول (١٥ - ١) .

الجدول (١٥ - ١)

عدد التوقعات س	التكرار ك	عدد التوقعات س	التكرار ك
٤٢	٢٣	١٤	١
٤١	٤	١١	١
٣٦	١	١٠	١
٣٢	١	٩	١
٢٩	١	٧	٠
٢٧	٢	٦	٣
٢٣	١	٥	٢
١٩	١	٤	٠
١٦	٢	٣	٠
١٥	٢		

$$\sum K = ٥٠ \quad \sum S = ٥٤٤٩٧$$

١ : واحد تقسيم بعدد كلى توقعات على هذا الاقمار  
٢ : واحد تقسيم بعدد كلى توقعات



الحل :

حجم المجتمع  $h = ٦٧٦$  بطاقة ، حجم العينة  $n = ٥٠$  بطاقة

$$\bar{s} \text{ (أولا) } = \frac{1}{n} \sum k s = \frac{1}{50} \times ١٤٧١ = ٢٩,٤٢$$

وهذا هو الوسط الحسابي لعدد التوقعات في العينة .

من (١) ، نقدر العدد الكلي للتوقعات بالمقدار

$$٢ = h \bar{s} = ٦٧٦ \times ٢٩,٤٢ = ١٩٨٨٨ \text{ توقعات .}$$

( ثانيا ) لإيجاد فترة الثقة المطلوبة نحتاج إلى إيجاد الخطأ المعياري للمجمامع وهذا

بدوره يحتاج إلى إيجاد تباين العينة  $s^2$  كالآتي .

$$s^2 = \frac{1}{n-1} \left[ \sum k^2 s - \frac{(\sum k s)^2}{n} \right]$$

$$= \frac{1}{49} \left( ٥٤٤٩٧ - \frac{(١٤٧١)^2}{50} \right) = ٢٢٠,٥٠٠$$

$$s = ١٤,٨٣$$

$$\text{من (٣) : } \frac{s}{\sqrt{n}} = \frac{١٤,٨٣}{\sqrt{50}} = ٢,١٠$$

$$١٣٥٠,٢١٠ =$$

بالرغم من أن التوزيع الأصلي لعدد التوقعات يبدو بعيدا عن الاعتدال إلا أننا نستطيع أن نعتبر أن توزيع المعاينة للمتوسطات ، وبالتالي للمجاميع ، هو توزيع معتدل على وجه التقريب لأن حجم العينة كبيرا ( $n = 50$ ) . ومن جدول المساحات أسفل المنحنى المعتدل المعيارى نجد أن القيمتين الحرجتين اللتين يقع بينهما ٨٠٪ من التوزيع هما  $\pm 1.28$  . وكما في البند (٦ - ٦ - ٢) أو البند (٦ - ١٠ - أولاً) نجد ما يلي :

الحد الأدنى لمجموع التوقعات  $= 19888 - 1391,48 \times 1.28 = 1810.7$   
الحد الأعلى لمجموع التوقعات  $= 19888 + 1391,48 \times 1.28 = 21669$   
وإذن الفترة (١٨١٠٧ ، ٢١٦٦٩) هي فترة ثقة بدرجة ٨٠٪ لعدد التوقعات على الائتماس .

( أظهر العد الكلى لجميع البطاقات أن عدد التوقعات ٢١٠٤٥ )

#### (١٥ - ٣ - ٢) تقدير النسبة ح ومجموع الوحدات :

في المجتمع ذى الحدين تكون كل وحدة من وحدات المجتمع منتمية إلى واحد من اثنين من الأقسام ١ ، أو ينصب اهتمامنا على تقدير الدليل ح وهو نسبة الوحدات التى تقع فى أحد القسمين وليكن القسم ١ وعلى تقدير المجموع الكلى للوحدات فى هذا القسم .

نفرض أننا أخذنا من هذا المجتمع عينة عشوائية بسيطة حجمها  $n$  . نذكر أنه فى توزيع ذى الحدين للعينات التى من الحجم  $n$  يكون للمتغير  $x$  وسط حسابى  $n$  ح وتباين  $n$  ح (١ - ح) كما يكون لنسبة هذا المتغير وسط حسابى ح وتباين  $\frac{ح(١ - ح)}{n}$  وعلى ذلك فإن تناول هذه الحالة يمكن أن يتخذ نفس طريقة تناول

المجتمع "كمى مع وضع ح بدلا من  $\mu$  ، ح (١ - ح) بدلا من  $\sigma$  .

وينتج ما يلي :

$$(٤) \quad \text{أولاً) النسبة } r = \frac{c}{n}$$

وهي نسبة عدد وحدات العينة التي تنتمي إلى القسم أ إلى العدد الكلي للوحدات في العينة ، هي تقدير غير متحيز للنسبة ح .

$$(٥) \quad \text{كما أن العدد } c = h \cdot r$$

هو تقدير غير متحيز لمجموع الوحدات الواقعة في القسم أ في المجتمع .

$$(٦) \quad \text{(ثانياً) المقدار } E_r = \frac{\frac{c}{h} - 1}{\frac{c}{h}} \sqrt{\frac{r(1-r)}{n}} =$$

هو تقدير للخطأ المعياري للنسبة ر

$$(٧) \quad \text{والمقدار } E_r = h \cdot E_r$$

هو تقدير للخطأ المعياري لمجموع الوحدات في القسم أ

والصيغ (٥) ، (٦) ، (٧) هي نفس الصيغ (١) ، (٢) ، (٣) بعد وضع ر بدلا من r ووضع ر (١ - ر) بدلا من r<sup>٢</sup> .

مثال (١٥ - ٢) :

في قائمة من ٣٠٤٢ اسما وعنوانا سحبت عينة عشوائية بسيطة من ٢٠٠ اسم فظهر فيها أن هناك خطأ في ٣٨ عنوانا . قدر العدد الكلي للعناوين التي تحتاج إلى تصحيح وأوجد الخطأ المعياري لهذا التقدير .

الحل :

حجم المجتمع ه = ٣٠٤٢ شخصا وحجم العينة ه = ٢٠٠ شخصا

$$r = \frac{38}{200} = 0,19 \quad (\text{نسبة العناوين الخاطئة في العينة})$$

من (٥) ، نقدر المجموع الكلي للعناوين الخاطئة بالمقدار

$$c = h \cdot r = 0,19 \times 3042 = 578 \text{ عنوانا خاطئا .}$$

من (٧) ، الخطأ المعياري لهذا المجموع هو :

$$٨٤,٣٨ = \frac{٠,٨١ \times ٠,١٩}{٢٠٠} \sqrt{٣٠,٤٢} = \frac{ع}{م}$$

وقد أهملنا عامل التصحيح لأن  $\frac{٢٠٠}{٣٠,٤٢} = \frac{٥}{٥}$  وهي نسبة صغيرة .

### (١٥ - ٣ - ٣) حجم العينة :

كما سبق القول مرارا ، كلما كبر حجم العينة كلما زادت ثقتنا فيما نستخلصه من نتائج . ولذلك ينبغي أن نحرص على ألا يكون حجم العينة صغيرا بدرجة تكون معها دقة تقديراتنا أقل مما يجب . غير أنه ينبغي في الوقت نفسه أن نتجنب أن يكون حجم العينة كبيرا بدرجة تثقل كاهلنا بالجهد والتكاليف . وبالتالي فإن الخطوة الأولى في عملية التجريب هي تحديد الحجم المناسب للعينة . وفي هذا الصدد نحيل القارئ إلى البند (٦ - ١٠) وبصفة خاصة إلى الصيغة (٢٣) التي تعطينا الحد الأعلى لحجم العينة عندما تكون المعاينة من مجتمع معتدل ، وهي :

$$(٨) \quad \left( \frac{\sigma^2}{\frac{١}{٢} \frac{ع}{خ}} \right) = ٥$$

والصيغتين (٢٦) ، (٢٧) في حالة المعاينة من مجتمع ذي حدين وهما :

$$(٩) \quad \text{حيث } ١ - ح = \frac{ع}{خ} \quad \left( \frac{\sigma^2}{\frac{١}{٢} \frac{ع}{خ}} \right) = ٥$$

$$(١٠) \quad \left( \frac{\sigma^2}{\frac{١}{٢} \frac{ع}{خ}} \right) = \frac{١}{٤} = ٥$$

فى كثير من الأحيان يعرف الباحث أو يكون لديه ما يدعو إلى الشك فى أن المجتمع غير متجانس من حيث المتغير الذى يدرسه ، بل ينقسم إلى عدد من القطاعات تختلف الاستجابات فيها بين كل قطاع وآخر بينما تتجانس داخل كل قطاع على حدة . وإذا كان الأمر كذلك نقول إن المجتمع مقسم إلى طبقات أو شرائح تحدها تركيبة المجتمع ، وهذه الطبقات قد تكون بحسب الجنس أو العمر أو الجنسية أو المستوى الثقافى أو درجة الإصابة بمرض ما . فى هذه الحال لا تكون العينة العشوائية البسيطة صالحة لتمثيل المجتمع ، بل تكون خطة المعاينة المناسبة هى تلك المسماة بالمعاينة العشوائية الطبقة البسيطة ، أو اختصارا بالمعاينة الطبقة . وتلخص هذه الخطة فى تحديد طبقات المجتمع بحيث لا تتداخل طبقة مع أخرى ثم أخذ عينة عشوائية بسيطة من كل طبقة بحيث تكون المعاينة مستقلة من طبقة إلى أخرى . وتتألف العينة المطلوبة من مجموع هذه العينات الجزئية .

وتبدأ الخطة بتحديد الحجم الكلى للعينة أو النسبة التى يرى أخذها من الحجم الكلى للمجتمع ، ثم تحديد أحجام العينات الجزئية مع الأخذ فى الاعتبار أحجام الطبقات والتباين داخل كل طبقة ، أو أى عوامل أخرى تؤثر فى تركيب المجتمع .

مثال (١٥ - ٣) :

نفرض أن لدينا مجتمعا حجمه ٤٠٠٠ وأن الإمكانات لا تسمح إلا بفحص عينة حجمها ٦٠ أى بنسبة  $\frac{60}{4000} = 0.015$  من حجم المجتمع . ونفرض أننا

نعرف أن المجتمع مقسم إلى ثلاث طبقات أحجامها ٨٠٠ ، ١٢٠٠ ، ٢٠٠٠ . نظرا لاختلاف أحجام الطبقات فإن العينة تكون أقدر تمثيلا للمجتمع إذا أخذنا للطبقة ذات الحجم الأكبر أن تسهم بقدر أكبر فى العينة ، وللطبقة ذات الحجم الأصغر أن تسهم بقدر أقل . ولتحقيق هذه العدالة نستخدم الطريقة الآتية .

## طريقة التقسيم المتناسب

### PROPORTIONAL ALLOCATION METHOD

تنص هذه الطريقة على أن نأخذ من كل طبقة عينة عشوائية بسيطة يتناسب حجمها مع حجم الطبقة . فإذا رمزنا لحجم المجتمع بالرمز  $H$  وللحجم الكلي للعينة بالرمز  $h$  وكان المجتمع مقسما إلى  $k$  من الطبقات أحجامها  $H_1, H_2, \dots, H_k$  ،  $h_1, h_2, \dots, h_k$  (مجموعها  $h$ ) فإننا نأخذ من هذه الطبقات عينات أحجامها  $h_1, h_2, \dots, h_k$  (مجموعها  $h$ ) بحيث

$$\frac{h_1}{H_1} = \dots = \frac{h_2}{H_2} = \frac{h_k}{H_k}$$

مع ملاحظة أن كلا من هذه النسب يساوى  $\frac{h}{H}$  وهو في هذا المثال يساوى  $0,015$  . ومن السهل أن نرى أن الحجم  $h_i$  الذي يؤخذ من الطبقة  $i$  يكون على الصورة

$$h_i = \frac{h}{H} \times H_i \quad \text{و} \quad 1, 2, \dots, k \quad (11)$$

ففي المثال (١٥ - ٣) تكون أحجام العينات الجزئية كما يلي - انظر الجدول (١٥ - ٢) :

$$h_1 = 2000 \times 0,015 = 30$$

$$h_2 = 1200 \times 0,015 = 18$$

$$h_3 = 800 \times 0,015 = 12$$

الجدول (١٥ - ٢)

أحجام العينات بطريقة التقسيم المتساوي

حجم العينة $n$	حجم الطبقة $n_h$	الطبقة
٣٠	٢٠٠٠	(١)
١٨	١٢٠٠	(٢)
١٢	٨٠٠	(٣)
٦٠	٤٠٠٠	المجموع

الجدول (١٥ - ٣)

أحجام العينات بطريقة التقسيم الأثقل

حجم العينة $n$	الانحراف المعياري $\sigma$	حجم الطبقة $n_h$	الطبقة
٣١	٤	٢٠٠٠	(١)
١٤	٣	١٢٠٠	(٢)
١٥	٥	٨٠٠	(٣)
٦٠		٤٠٠٠	المجموع

إن المعاينة بطريقة التقسيم المتناسب تأخذ في الاعتبار الفروق بين أحجام الطبقات ولا تأخذ في الاعتبار الفروق بين التباينات داخل هذه الطبقات بل تعتبر أن هذه التباينات متساوية . وإذا كانت التباينات تختلف من طبقة لأخرى فمن الأفضل أن نأخذ عينات أكبر حجما من الطبقات الأكثر تشتتا وعينات أصغر حجما من الطبقات الأقل تشتتا ، ولتحقيق ذلك نستخدم الطريقة الآتية .

### طريقة التقسيم الأمثل OPTIMUM ALLOCATION METHOD

تنص هذه الطريقة على أن نأخذ من كل طبقة عينة عشوائية بسيطة يتناسب حجمها مع كل من حجم الطبقة وانحرافها المعياري ، فإذا كان  $\sigma_1, \sigma_2, \dots$  ،  $\sigma_k$  ترمز إلى الانحرافات المعيارية للطبقات فإن الأحجام التي تؤخذ من هذه الطبقات تكون بحيث :

$$\frac{n_1}{\sigma_1} = \dots = \frac{n_2}{\sigma_2} = \frac{n_k}{\sigma_k}$$

ويمكن إثبات أن هذه المتساويات تؤدي إلى أن يكون الحجم  $n_i$  الذي يؤخذ من الطبقة  $i$  على الصورة الآتية :

$$n_i = \frac{n \times \sigma_i}{\sigma_1 + \sigma_2 + \dots + \sigma_k} \quad i = 1, 2, \dots, k$$

$$(12) \quad n_i = \frac{n \times \sigma_i}{\sigma_1 + \sigma_2 + \dots + \sigma_k}$$

ويلاحظ أنه إذا كانت الانحرافات المعيارية للطبقات متساوية جميعها فإن الصيغة (12) تتحول بالضبط إلى الصيغة (11) .



وفي المثال إذا كان  $\sigma_1 = 4$  ،  $\sigma_2 = 3$  ،  $\sigma_3 = 5$  فإن أحجام العينات الجزئية تحسب كما يلي : انظر الجدول (١٥ - ٣) .

$$\sigma_{\text{مجموع}} = 2000 \times 4 + 1200 \times 3 + 800 \times 5 = 15600$$

$$n_1 = \frac{4 \times 2000}{15600} = 31$$

$$n_2 = \frac{3 \times 1200}{15600} = 14$$

$$n_3 = \frac{5 \times 800}{15600} = 15$$

هذا مع ملاحظة أنه عند التعويض في أى من الصيغتين (١١) أو (١٢) نأخذ أقرب عدد صحيح للقيمة التي تنتج من هذا التعويض . وفي الصيغة (١٢) إذا كانت الانحرافات المعيارية للطبقات غير معروفة فينبى تقديرها من عينات سابقة .  
تقدير البارامترات :

بعد تحديد أحجام العينات سواء بطريقة التقسيم المناسب أو التقسيم الأمثل نقوم بسحب العينات من الطبقات بحسب هذه الأحجام ثم نجري ما نريد من قياسات كقياس الطول أو الوزن ... على وحدات هذه العينات لنحصل على مجموعة من القيم لكل عينة . من هذه القيم نحسب متوسطات العينات  $\bar{y}_1$  ،  $\bar{y}_2$  ، ... ،  $\bar{y}_r$  وانحرافات المعيارية  $s_1$  ،  $s_2$  ، ... ،  $s_r$  وذلك لاستخدامها فيما يلي :

( أولاً ) تقدير متوسطات الطبقات :

نظرا لأن العينات المسحوبة هي عينات عشوائية بسيطة فإن متوسط الطبقة ومجموعها والخطأ المعياري لتوزيع المعاينة تقدر بنفس الصيغ المبينة بالبند (١٥ - ٣ - ١) السابق .

(ثانياً) تقدير متوسط المجتمع والخطأ المعياري .

(أ) تقدير الوسط الحسابي

يقدر الوسط الحسابي  $\mu$  للمجتمع تقديراً غير متحيز بالمقدار  $\bar{y}$  حيث

$$\bar{y} = \frac{1}{n} (y_1 + y_2 + \dots + y_n)$$

$$(13) \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

مع ملاحظة أن متوسطات العينات  $\bar{y}_i$  قد رجحت بأحجام الطبقات وليس بأحجام العينات .

وحيث تكون المعاينة بطريقة التقسيم المتناسب فإن  $\frac{y_i}{n} = \frac{y_i}{n_i} \cdot \frac{n_i}{n}$

وذلك من الصيغة (١١) وفي هذه الحالة تتحول الصيغة (١٣) إلى الصيغة الآتية :

$$(14) \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \cdot \frac{n_i}{n}$$

أي أن الوسط الحسابي  $\mu$  للمجتمع يقدر في هذه الحالة بواسطة الوسط الحسابي للملاحظات في العينة الكلية التي تنتج من ضم العينات الجزئية معا .  
وفي كلتا الحالتين يقدر المجموع الكلي للمجتمع بالمقدار .

$$(15) \quad \sum y_i = \sum y_i \cdot \frac{n_i}{n}$$

(ب) تقدير الخطأ المعياري .

من الصيغة (١٣) يمكن إثبات أن التباين  $\sigma^2$  لتوزيع المعاينة للوسط الحسابي يقدر بدلالة تباينات العينات  $\sigma_i^2$  بالمقدار  $\sigma^2$  حيث

$$(16) \quad \sigma^2 = \sum_{i=1}^n \sigma_i^2 \cdot \frac{n_i}{n} \cdot \left( \frac{n_i}{n} - 1 \right)$$

وحيث  $v = \frac{v}{n} =$  نسبة حجم الطبقة  $v$  إلى الحجم الكلي للمجتمع .

أما الخطأ المعياري للمتوسطات فيقدر بالجذر التربيعي لهذا المقدار .  
وإذا كانت المعاينة الطبقية بالتقسيم المتناسب للأحجام فإن عوامل التصحيح تكون واحدة لجميع الطبقات وكل منها يساوى  $1 - \frac{v}{n}$  كما أن

$$v \times \frac{v}{n} = v^2$$

وبالتعويض بهذا المقدار في (١٦) تؤول إلى الصيغة الآتية

$$e' = \frac{1}{n} (1 - \frac{v}{n}) e^2 \quad (١٧)$$

وإذا أمكن أن نعتبر أن تباينات الطبقات  $e^2$  متساوية وكل منها يساوى  $e^2$  فإن الصيغة (١٧) لحالة التقسيم المتناسب تؤول إلى الصيغة البسيطة الآتية :

$$e' = \frac{e^2}{n} (1 - \frac{v}{n}) \quad (١٨)$$

والجذر التربيعي لهذه الصيغة هو بالضبط الصيغة (٢) لتقدير الخطأ المعياري في حالة المعاينة العشوائية البسيطة من مجتمع متبني فيما عدا أن التباين المشترك  $e^2$  يحسب هنا من داخل العينات الجزئية كالآتي :

$$e^2 = \frac{1}{n} [e_1^2 (1 - \frac{v_1}{n}) + e_2^2 (1 - \frac{v_2}{n}) + \dots + e_k^2 (1 - \frac{v_k}{n})] \quad (١٩)$$

وفي جميع الحالات تقدر تباينات المجاميع من تباينات المتوسطات بالضرب في مربع حجم المجتمع وهو  $n$  .

مثال (١٥ - ٤) :

في المثال (١٥ - ٣) إذا كانت أحجام العينات قد حسبت بطريقة التقسيم المتناسب ، وعلى فرض تساوى تباينات الطبقات ، فابعد فترة ثقة بدرجة ٩٥٪ لتوسط المجتمع علما بأن الأوساط الحسابية والانحرافات المعيارية للعينات كالآتي :  
 $\bar{x}_1 = 7$  ،  $s_1^2 = 10$  ،  $\bar{x}_2 = 13$  ،  $s_2^2 = 4$  ،  $\bar{x}_3 = 2$  ،  $s_3^2 = 25$  ،  $\bar{x}_4 = 4$  .

الحل :

من الجدول (١٥ - ٢) والصيغة (١٤) نجد أن

$$s_p^2 = \frac{1}{n} \sum_{i=1}^k s_i^2$$

$$s_p^2 = \frac{1}{16} (13 \times 12 + 10 \times 18 + 7 \times 30) = 9,1$$

من الصيغة (١٩) نقدر التباين المشترك للطبقات كالآتي :

$$s_{p, \bar{x}}^2 = \frac{1}{3-6} (16 \times 11 + 6,25 \times 17 + 9 \times 29) = 9,53$$

$$\therefore s_{p, \bar{x}} = 3,09$$

من الصيغة (١٨) نقدر الخطأ المعياري لتوزيع المعاينة للأوساط الحسابية كالآتي مع ملاحظة أن عامل التصحيح قريب من الواحد ويمكن إهماله :

$$s_{\bar{x}} = \frac{s_{p, \bar{x}}}{\sqrt{n}} = \frac{3,09}{\sqrt{16}} = 0,77$$

ولما كان حجم العينة كبيرا ( $n = 16$ ) يمكن أن نعتبر أن توزيع المعاينة معتدلا ويكون حدا الثقة بدرجة ٩٥٪ لتوسط المجتمع على الصورة الآتية :

$$\bar{S} \pm E \times 1,96$$

$$\therefore \text{الحد الأدنى للفترة} = 9,1 - 0,40 \times 1,96 = 8,32$$

$$\text{والحد الأعلى للفترة} = 9,1 + 0,40 \times 1,96 = 9,88$$

وبذلك تكون الفترة (8,32 ، 9,88) هي فترة ثقة بدرجة 90٪ لمتوسط المجتمع .

مثال (١٥ - ٥) :

في المثال (١٥ - ١) إذا كانت أحجام العينات قد حسبت بطريقة التقسيم الأمثل فابعد فترة ثقة بدرجة 99٪ لمتوسط المجتمع علماً بأن الأوساط الحسابية للعينات هي  $\bar{S}_1 = 8$  ،  $\bar{S}_2 = 12$  ،  $\bar{S}_3 = 13$  وأن الانحرافات المعيارية للطبقات هي كما جاءت بالجدول (١٥ - ٣) :  $\sigma_1 = 4$  ،  $\sigma_2 = 3$  ،  $\sigma_3 = 5$  .

الحل :

من الجدول (١٥ - ٣) والصيغة (١٣) نجد أن

$$\bar{S} = \frac{1}{N} \sum S_i$$

$$= \frac{1}{4000} (8 \times 2000 + 12 \times 1200 + 13 \times 800) = 10,2$$

من الصيغة (١٦) نقدر الخطأ المعياري لتوزيع المعاينة للأوساط الحسابية كالتالي ، مع إهمال عوامل التصحيح لقرب كل منها من الواحد الصحيح .

$$\text{لدينا : } \sigma_{\bar{S}} = \sqrt{\frac{2000}{4000}} = 0,707 \text{ ، } \sigma_{\bar{S}} = \sqrt{\frac{1200}{4000}} = 0,547 \text{ ، } \sigma_{\bar{S}} = \sqrt{\frac{800}{4000}} = 0,447$$

$$٠,٠٤ = {}^1_r \left( \frac{٨٠٠}{٤٠٠٠} \right) = {}^1_r ,$$

$$٠,٢٥٤ = \frac{٢٥ \times ٠,٠٤}{١٥} + \frac{٩ \times ٠,٠٩}{١٤} + \frac{١٦ \times ٠,٢٥}{٣١} = {}^٢_c .$$

$$٠,٥٠ = {}^٣_c .$$

ويكون هذا الثقة بدرجة ٩٩٪ لتوسط المجتمع على الصورة

$$\bar{x} \pm ٢,٥٨ \times {}^٣_c$$

وبالتعويض في هذه الصيغة نجد أن الفترة (٨,٩١ ، ١١,٤٩) هي فترة ثقة بدرجة ٩٩٪ لتوسط المجتمع .

مثال (١٥ - ٦) :

حسب تعداد السكان في سنة ما في ٦٤ مدينة من مدن إحدى الدول . وقد قسمت هذه المدن إلى طبقتين تتألف الأولى من المدن الأكبر حجما وعددها ١٦ مدينة وتتألف الثانية من الـ ٤٨ مدينة الباقية ، ولخصت البيانات في الجدول الآتي .

الجدول (١٥ - ٤)

الطبقة	الحجم $n_i$	$\bar{x}_i$	$s_i^2$
(١)	١٦	١٠٠٧٠	٧١٤٥٥٤٠
(٢)	٤٨	٩٤٩٨	٢١٤١٧٢٠

إذا قُدر المجموع الكلي للسكان في تلك السنة من عينة من ٢٤ مدينة فأوجد الخطأ المعياري لهذا التقدير

(أولاً) إذا كانت العينة عشوائية بسيطة ،

(ثانياً) إذا كانت العينة طبقية ذات تقسيم متناسب ،

(ثالثاً) إذا كانت العينة طبقية وأخذت ١٢ مدينة من كل طبقة .

الحل :

من البيانات المعطاة نستطيع أن نحسب تباين المجتمع وتباينات الطبقات ولا حاجة لنا إذن لتقدير هذه التباينات من العينة .

(أولاً) إذا كانت العينة عشوائية بسيطة فإنها تكون قد أخذت من المجتمع ككل بصرف النظر عن الطبقات . ويكون لدينا ما يلي :

حجم المجتمع  $N = ٦٤$  مدينة ، حجم العينة  $n = ٢٤$  مدينة

$\Sigma x = ١٠٠٧٠ + ٩٤٩٨ = ١٩٥٦٨$  ألف نسمة (مجموع السكان)

$\Sigma x^2 = ٧١٤٥٥٤٠ + ٢١٤١٧٢٠ = ٩٢٨٧٢٦٠$

$\therefore$  تباين المجتمع  $\sigma^2 = \frac{1}{64} [ \frac{(19568)^2}{24} - 9287260 ] = ٥١٦٢٨,٩٧$

$\therefore \sigma = ٢٢٧,٢٢$

من الصيغة (٣) :  $\frac{\sigma}{\sqrt{n}} = \frac{227,22}{\sqrt{24}} = \frac{\sqrt{227,22 \times 64}}{24} = \sqrt{\frac{227,22 \times 64}{24}}$

$= ٢٣٤٦,٧٢$  (الخطأ المعياري لمجموع السكان)

(ثانيا) إذا كانت المعالجة طبقية وبالتقسيم المتناسب فإن حجمي العييتين يكونان كالآتي :

$$18 = 48 \times \frac{24}{72} = 12 , 6 = 16 \times \frac{24}{72} = 4$$

$$0.477,74 = \left[ \frac{(10070)}{16} - 7140400 \right] \frac{1}{16} = 1^{\circ} \text{ع}$$

$$0.464,60 = \left[ \frac{(9498)}{48} - 2141720 \right] \frac{1}{48} = 2^{\circ} \text{ع}$$

نلاحظ أن تباين الطبقة الأولى حوالى عشرة أمثال تباين الطبقة الثانية ولذلك لا نستطيع اعتبارهما متساويين .

بضرب الصيغة (١٧) فى مربع حجم المجتمع ينتج أن تباين مجموع المجتمع هو

$$[0.464,60 \times 48 + 0.477,74 \times 16] \left( \frac{24}{72} - 1 \right) \frac{24}{72 \times 24} = 1^{\circ} \text{ع}$$

$$1783240 = 10.69947 \times \frac{24}{72} =$$

$$1330,28 = 1^{\circ} \text{ع} \therefore$$

(ثالثا) إذا كانت المعالجة طبقية وأخذنا  $n_1 = 12$  ،  $n_2 = 12$  نستخدم الصيغة العامة (١٦) بعد ضربها فى مربع حجم العينة مع ملاحظة ما يلى :

$$12 = 48 \times \frac{24}{72} = 12 , 12 = 16 \times \frac{24}{72} = 12$$

$$\frac{24}{48} = \frac{12}{48} - 1 , \frac{24}{16} = \left( \frac{12}{16} - 1 \right) ,$$

$$\left[ \frac{36}{48} \times 0.464,60 \times \frac{24}{72} + \frac{24}{16} \times 0.477,74 \times \frac{12}{72} \right] \frac{24}{12} = 1^{\circ} \text{ع}$$



$$\frac{1}{11} = (36 \times 0.464, 60 \times 48 + 4 \times 0.477, 74 \times 16)$$

$$1.06124, 1 =$$

$$1.27, 68 = \epsilon \therefore$$

في هذا المثال ، بمقارنة الأخطاء المعيارية وهي ٢٣٤٦,٧٢ ، ١٣٣٥,٣٨ ، ١٠٢٧,٦٨ نجد أن أخذ حجمين متساويين للعنتين كان أكثر دقة من طريقة التقسيم المتناسب وكلاهما أدق كثيرا من طريقة المعاينة العشوائية البسيطة .

### المعاينة الطبقية من مجتمع ذي حدين :

نستخدم نفس الصيغ التي قدمت في حالة المعاينة الطبقية من مجتمعات كمية مع وضع النسب  $p_1, p_2, \dots, p_r$  المحسوبة من العينات بدلا من المتوسطات الحسابية ووضع التباينات  $s_1^2, s_2^2, \dots, s_r^2$  بدلا من التباينات  $\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2$  . ففي تحديد أحجام العينات نستخدم نفس الصيغة (١١) وهي

$$(20) \quad n_h = \frac{N}{h} \times p_h$$

في حالة استخدام طريقة التقسيم المتناسب . أما في حالة التقسيم الأمثل فنستخدم الصيغة (١٢) بعد وضعها كالتالي :

$$(21) \quad n_h = \frac{N}{h} \times \frac{p_h}{\sqrt{(1-p_h)}}$$

حيث  $p_h$  هو نسبة وقوع الحدث في الطبقة  $h$  .

وفي تقدير النسبة  $p_h$  وهي احتمال وقوع الحدث في المجتمع نستخدم الصيغة (١٣) بعد وضعها في الصورة الآتية :

$$(٢٢) \quad s = \frac{1}{2} \sqrt{\frac{L}{H}} \quad s$$

وحيث تكون المعاينة بطريقة التقسيم المتناسب تؤول هذه الصيغة إلى :

$$(٢٣) \quad s = \frac{1}{n} \sqrt{\frac{L}{H}} \quad s$$

وهذا يعنى أننا في هذه الحالة نقدر الاحتمال ح في المجتمع بواسطة النسبة المشاهدة في العينة الكلية التي تتألف من ضم جميع العينات الجزئية .

كذلك ، لتقدير الخطأ المعياري  $\sigma$  لتوزيع المعاينة للنسبة ر نستخدم الصيغة (١٦) بعد وضعها كالآتي :

$$(٢٤) \quad \sigma = \sqrt{\frac{L}{H} \times \frac{s(s-1)}{n} \left( \frac{n}{H} - 1 \right)}$$

حيث  $s = H / n$

وحيث تكون المعاينة بطريقة التقسيم المتناسب وعلى فرض تساوى تباينات الطبقات نستخدم الصيغة (١٨) وهي

$$(٢٥) \quad \sigma = \frac{1}{\sqrt{H}} \sqrt{\frac{L}{H} - 1}$$

$$(٢٦) \quad \text{حيث } \sigma' = \frac{1}{\sqrt{H} - 1} \times (1 - \frac{n}{H}) \times s (s - 1)$$

هو تقدير للتباين المشترك للطبقات .

مثال (١٥ - ٧) :

كان عدد الأطفال في إحدى المدن الكبرى ٤٣١٥٤٢ طفلاً . وفي إحدى التجارب كان المطلوب تحديد فترة ثقة بدرجة ٩٥٪ لنسبة الإصابة بمرض ما بين هؤلاء الأطفال عن طريق عينة من ٤٥٠ طفلاً أى بواقع  $\frac{450}{431542} = 0.00104$  تقريباً . وقد قسم المجتمع إلى ثلاث طبقات بحسب كثافة السكان ، إذ كان من المتقَد أن الإصابة بهذا المرض تختلف باختلاف هذه الكثافة . وقد وجد أن أعداد الأطفال في هذه الطبقات ٢٢٠٣٨٦ ، ١٤٦٠٣١ ، ٦٥١٢٥ .

الحل :

نظراً لعدم وجود معلومات عن تباينات الإصابة بالمرض في الطبقات فقد استخدم لتحديد أحجام العينات طريقة التقسيم المتناسب بحسب الصيغة (١١) وهي

$$n_i = n \times \frac{N_i}{N}$$

$$\therefore n_1 = \frac{450}{431542} \times 220386 = 230 \text{ طفلاً}$$

$$n_2 = \frac{450}{431542} \times 146031 = 152 \text{ طفلاً}$$

$$n_3 = \frac{450}{431542} \times 65125 = 68 \text{ طفلاً}$$

وقد اختيرت عينات عشوائية بسيطة من تلك الطبقات بحسب الأحجام الناتجة . ويفحص الأطفال وجد أن أعداد الأطفال المصابين بالمرض ونسبة هذه الإصابة في العينات كما هو مبين بالعمودين الآخرين من الجدول (١٥ - ٥) الآتي :

المجدول (١٥ - ٥)

أعداد المصابين بالمرض ونسب هذه الإصابة

نوع الطبقة	حجم الطبقة هـ	حجم العينة ن	عدد المصابين في العينة	نسبة الإصابة % س
شديدة الأزدحام	٢٢٠٣٨٦	٢٣٠	١٦٤	٧١,٣
متوسطة الأزدحام	١٤٦٠٣١	١٥٢	٩٢	٦٠,٥
قليلة الأزدحام	٦٥١٢٥	٦٨	١٨	٢٦,٥
المجموع	٤٣١٥٣٢	٤٥٠	٢٧٤	٦٠,٩

نظراً لأننا استخدمنا طريقة التقسيم المتناسب فإن نسبة الإصابة بالمرض في مجتمع الأطفال تقدر بنسبة الإصابة في العينة الكلية وهي

$$r = ٦٠,٩\%$$

نظراً لافتراضنا أن التباينات متساوية في الطبقات فإننا نقدر التباين المشترك بالصيغة (٢٦) كما يلي :

$$s^2 = \frac{1}{447} (٢٢٩ \times ٠,٧١٣ \times ٠,٢٨٧ + ١٥١ \times ٠,٦٠٥ \times ٠,٣٩٥)$$

$$+ ٦٧ \times ٠,٢٦٥ \times ٠,٧٣٥)$$

$$= ٠,٢١٤٨$$

$$\therefore \text{ع} = ٠,٤٦٣$$

∴ الخطأ المعياري لتوزيع المعاينة للنسبة  $\pi$  للعينات التي من الحجم ٤٥٠ هو حسب الصيغة (٢٥) :

$$\text{ع} = \frac{٠,٤٦٣}{\sqrt{٤٥٠}} = ٠,٠٢٢ \text{ (مع إهمال عامل التصحيح لقربه من الواحد)}$$

في عينة بالحجم ٤٥٠ يكون المتوسط النسبي موزعا توزيعا معتدلا على وجه التقريب وبالتالي يكون حد الثقة بدرجة ٩٥٪ لمتوسط الإصابة بالمرض في المجتمع  $\pi$  ما  $\pi \pm \text{ع}$   $١,٩٦$  أى  $٠,٦٠٩ \pm ٠,٠٢٢ \times ١,٩٦$  . وبحساب هاتين القيمتين نجد أن الفترة المطلوبة هي (٥٦,٦ ٪ ، ٦٥,٢ ٪) .

## MULTISTAGE SAMPLE العينة المتعددة المراحل (١٥ - ٥)

(أو العينة العشبية nested sample)

حين يكون المجتمع كبيرا نضطر أحيانا إلى اختيار العينة عن طريق سلسلة من المراحل . وكمثال لذلك نفرض أننا نريد اختيار عينة لتقدير عدد الحالات من المرضى الذين فحصوا بالأشعة في أسبوع في المستشفيات الحكومية بدولة ما . في هذه الحال يصعب بل يستحيل تصميم خطة للمعاينة من المرضى مباشرة ، ولذلك نلجأ إلى المعاينة على مراحل كما يلي . نجرى حصرا بالمحافظات أو المناطق الجغرافية التي بها مستشفيات حكومية . تبدأ المعاينة باختيار عينة عشوائية بسيطة من هذه المناطق حجمها  $\pi$  . منطقة ونسجل أسماء المستشفيات الحكومية بكل منها ، وهذه هي المرحلة الأولى . نأخذ من كل منطقة من المناطق السابق اختيارها عينة عشوائية بسيطة من المستشفيات حجمها  $\pi$  من المستشفيات وهذه هي المرحلة الثانية وبذلك يكون لدينا  $\pi \times \pi$  مستشفى . نأخذ من كل من هذه المستشفيات

عينة عشوائية بسيطة من المرضى الذين دخلوها أو كانوا مقيمين بها في الأسبوع المحدد وليكن حجمها  $n$  مريضاً وهذه هي المرحلة الثالثة والأخيرة ، وبذلك نكون قد حصلنا على عينة إجمالية حجمها  $n_1 \times n_2 \times n_3$  من المرضى ، ونستطيع حيثذ أن تفحص ملفاتهم لمعرفة عدد الذين فحصوا بالأشعة .

وقد يكون من المناسب أحيانا استخدام التقسيم الطبقي في واحدة أو أكثر من مراحل المعاينة إذا استدعى الأمر ذلك فتقسم المناطق الجغرافية مثلاً إلى مدن كبيرة ومدن صغيرة وقرى ، أو تقسم للمستشفيات بحسب التخصص ، أو يقسم المرضى بحسب الجنس .

وكمثال آخر ، نفرض أننا نريد تقدير متوسط طول فتلة القطن في بالة كبيرة من القطن . نأخذ عدة حفنات من القطن عشوائيا من جوانب مختلفة من البالة وهذه مرحلة أولى . نأخذ كل حفنة من الحفنات التي اخترناها ونقسمها إلى جزعين نرمي أحدهما ونحتفظ بالآخر وهذه مرحلة ثانية . نكرر هذه العملية عدة مرات حتى نحصل على عدد مناسب من الفتلات لقياسها وحساب متوسط الطول فيها .

يلاحظ في هذا المثال أن الوحدات المختارة في كل مرحلة على هيئة « مجموعات » aggregates وليست على هيئة مفردات . في مثل هذه الحالة توصف المعاينة بأنها معاينة عنقودية cluster sampling ومن أمثلتها أيضا سحب عينات من رمال أحد الشواطئ أو سحب عبوات من ماء يجري نهر ، أو سحب فصول كاملة من عدد من المدارس .

ومن الحالات التي تستلزم المعاينة المتعددة المراحل تلك التي تحتاج إلى إجراء الاختبارات الكيميائية أو الفيزيائية أو البيولوجية التي يصعب إجراؤها إلا على أجزاء صغيرة من المادة المختبرة كما في المثال (١٥ - ٨) الآتي .

## التحليل الإحصائي :

يحتاج تحليل العينات متعددة المراحل إلى استخدام أسلوب تحليل التباين بالتموج عشوائى التأثيرات - راجع البند (٨ - ١٤) - ولترى ذلك نبدأ بتناول العينة ذات المرحلتين مستعنيين بالمثال (٨ - ١٦) حيث كان اهتمامنا بتقدير نسبة الكلسيوم في أوراق اللفت الأخضر وكانت المعاينة على مرحلتين أولهما أخذ عينة عشوائية من أوراق النبات ، وتسمى وحدات هذه العينة بالوحدات الابتدائية primary sampling units ، وثانيهما أخذ عينة عشوائية من أجزاء كل ورقة وقياس نسبة الكلسيوم ، وتسمى وحدات هذه العينة بوحدات المرحلة الثانية second-stage units, or sub-units . ولقد اعتبرنا أن الوحدات الابتدائية هي عينة عشوائية من « المعالجات » وأن نسب الكلسيوم هي القيم الناتجة تحت تأثير هذه المعالجات . ومن ثم كان تحليلنا للبيانات عن طريق تحليل التباين بالتموج عشوائى التأثيرات الذى يأخذ الصيغة الآتية :

$$S_{\text{مجموع}} = \mu + \alpha + \beta + \gamma \quad (27)$$

لقد قدمنا في البند (٨ - ١٤) تفصيلا لهذا التحليل مدعما بالمثالين (٨ - ١٦) و(٨ - ١٧) وليس هناك ما يدعو لتكرار ذلك هنا .

نقوم الآن بتحليل عينة ذات ثلاث مراحل ونستعين في ذلك بالمثال (٨ - ١٥) الآتى .

مثال (٨ - ١٥) :

اعتبر تجربة المثال (٨ - ١٦) عن محتوى الكلسيوم في أوراق نبات اللفت الأخضر . وافرض أننا لم نبدأ باختيار عينة عشوائية من الأوراق بل بدأنا بعينة عشوائية من نبات اللفت ذاته حجمها  $k = 4$  نباتات (الوحدات الابتدائية) ثم أخذنا من كل نبات عينة عشوائية من  $l = 3$  ورقات (وحدات المرحلة الثانية)

ثم أخذنا من كل ورقة عينة عشوائية من  $b = 2$  من الأجزاء وزن كل منها ١٠٠ ملليجرام ( وحدات المرحلة الثالثة ) فحصلنا بذلك على  $4 = 2 \times 2$  جزءا من أوراق النبات هي التي نقوم بقياس محتوى الكلسيوم فيها . نفرض أن القياسات جاءت كما في الجدول (١٥ - ٦) الآتي .

الجدول (١٥ - ٦)

النسب المئوية للكلسيوم في  $b = 2$  جزءا من كل من  $a = 2$  ورقة من كل من  $c = 4$  نباتا

	(١)		(٢)		(٣)		(٤)		النبات
	ب	ا	ب	ا	ب	ا	ب	ا	الأوراق
نسبة الكلسيوم	٢,٢٨	٢,٥٢	٢,٤٦	١,٨٧	٢,١٩	٢,٧٧	٢,٧٤	٢,٥٥	٢,٧٨
في جزء كل ورقة	٢,٨٠	٢,٤٨	٢,٤٥	١,٩٢	٢,١٩	٢,٦٦	٢,٤٤	٢,٥٥	٢,٨٧
المجموع للورقة	٥,٦٨	٢,٠٠	٤,٩٠	٣,٧٩	٤,٣٨	٥,٤٣	٧,١٨	٥,١٠	٨,١٩
المجموع لنبات	١٩,٠٥	١٣,٠٧	١٧,٧١	٢٢,٤٦	٧٢,٢٩				

( في هذا المثال أخذنا عددا متساويا من الأوراق من كل نبات وكان من الممكن أي يختلف هذا العدد من نبات إلى آخر . وكذلك بالنسبة لعدد الأجزاء التي أخذت من كل ورقة . )

إذا كانت  $\mu$  ترمز إلى نسبة الكلسيوم في الجزء  $a$  من الورقة  $b$  من النبات  $c$  فإن النموذج الإحصائي للتحليل هو امتداد للنموذج (٢٣) :

$$\mu + \alpha + \beta + \gamma + \epsilon = \text{مجموع}$$

$$(٢٨) \quad \mu = ٢,٤١ \text{ و } \alpha = ١,٨٧, \beta = ٢,٠٠, \gamma = ٢,٤٤, \epsilon = ٤,٣٨$$



حيث أن تشتمل إلى النباتات ، بـ  $\sigma$  تشتمل إلى الأوراق .  
ولإمكانية التحليل الإحصائي سنفترض كالمعتاد أن

أ : مع  $(\sigma, 0)$  ، بـ : مع  $(\sigma, 0)$  ، جـ : مع  $(\sigma, 0)$  (٢٩)

في تحليل التباين نفصل مجموع المربعات الكلي للقياسات (نسب محتوى الكلسيوم) إلى مصادر مستقلة للاختلاف . وفي هذا المثال نجد أن هذه المصادر هي : النباتات - أوراق نفس النبات - محتوى الكلسيوم بأجزاء نفس الورقة . ولايجاد الاختلافات في هذه المصادر نحسب بالطريقة المعتادة كلا من  $\sigma^2$  ( الكلي ) ،  $\sigma^2$  ( بين النباتات ) ،  $\sigma^2$  ( بين الأوراق ) أما الاختلافات الباقية فنحسبها من هذه الاختلافات كالآتي :

(١)  $\sigma^2$  ( بين أوراق نفس النبات ) =  $\sigma^2$  ( بين الأوراق ) -  $\sigma^2$  ( بين النباتات )

(٢)  $\sigma^2$  ( بين القياسات على نفس الورقة ) =  $\sigma^2$  ( الكلي ) -  $\sigma^2$  ( بين الأوراق )

وفي المثال نجد من الجدول (١٥ - ٦) ما يلي .

$$\text{عامل التصحيح} = \frac{\sigma^2}{n} = \frac{72,29}{24} = 3,012$$

$$\sigma^2 \text{ (الكلي)} = 3,28 + 3,09 + \dots + 3,31 + 3,31 = 217,7435$$

$$10,2704 = \text{بدرجات حرية } n - 1 = 23$$

$$\sigma^2 \text{ (بين النباتات)} = \frac{1}{4} (19,05 + 13,07 + 17,71 + 22,46) - 217,7435 =$$

$$7,063 = \text{بدرجات حرية } k - 1 = 3$$

$$٢٢ \text{ (بين الأوراق)} = \frac{١}{٧} (٦,٣٧ + ٧,٠٠ + \dots + ٨,١٩ + ٦,٦٢) - ٢١٧,٧٤٣٥ =$$

$$١٠,١٩٠٥ = \text{بدرجات حرية أ ك} - ١ = ١١$$

$$٢٢ \text{ (بين أوراق نفس النبات)} = ١٠,١٩٠٥ - ٧,٥٦٠٣ =$$

$$٢,٦٣٠٢ = \text{بدرجات حرية أ ك} - ٨ =$$

$$٢٢ \text{ (بين القياسات على نفس الورقة)} = ١٠,٢٧٠٤ - ١٠,١٩٠٥ =$$

$$٠,٠٧٩٩ = \text{بدرجات حرية ب أ ك} - ١٢ =$$

وينتج جدول التباين الآتي :

المجموع (١٥ - ٧)

مصدر التباين	٢٢	د ح	تقدير التباين	التباين المتوقع
بين النباتات	٧,٥٦٠٣	٣	ع <sup>٢</sup> ٢,٥٢٠١ =	$\sigma^2 + \sigma^2$ $\sigma^2 + \sigma^2$
بين الأوراق داخل النباتات	٢,٦٣٠٢	٨	ع <sup>٢</sup> ٠,٣٢٨٨	$\sigma^2 + \sigma^2$
بين القياسات داخل الأوراق داخل النباتات	٠,٠٧٩٩	١٢	ع <sup>٢</sup> ٠,٠٠٦٧	
الكل	١٠,٢٧٠٤	٢٣		

من العمود الأخير نلاحظ أن كل مركبة من مركبات التباين داخلية في المركبة السابقة لها ، ولهذا يمكن بسهولة أن نرى ما يلي :

$$ع' \text{ هو تقدير غير متحيز للتباين } \sigma^2 \quad (٣٠)$$

$$\frac{1}{b} (ع' - ع) \text{ هو تقدير غير متحيز للتباين } \sigma^2 \quad (٣١)$$

$$\frac{1}{a} (ع' - ع) \text{ هو تقدير غير متحيز للتباين } \sigma^2 \quad (٣٢)$$

ففى هذا المثال نجد التقديرات الآتية :

$$(١) \text{ تقدير } \sigma^2 = ٠,٠٠٦٧$$

$$(٢) \text{ تقدير } \sigma^2 = \frac{1}{4} = (٠,٠٠٦٧ - ٠,٣٢٨٨) = ٠,١٦١١$$

$$(٣) \text{ تقدير } \sigma^2 = \frac{1}{4} = (٠,٣٢٨٨ - ٢,٥٢٠١) = ٠,٣٦٥٢$$

كما نجد التقديرين الآتين :

(٤) يقدر الوسط الحسابى  $\mu$  للنسبة المئوية للكليسيوم فى المجتمع بالوسط الحسابى للعينه وهو :

$$\bar{x} = \frac{٧٢,٢٩}{٢٤} = ٣,٠١$$

(٥) يقدر الخطأ المعيارى لتوزيع المعاينة للوسط الحسابى كالاتى :

$$(٣٣) \quad \frac{\text{التباين بين النباتات (الوحدات الابتدائية)}}{\text{حجم العينة}} = \frac{ع'}{n} = \sigma^2 \text{ تقدير}$$

$$٠,١٠٥ = ٢٤ \div ٢,٥٢٠١ =$$

ومن الجذر التربيعى لهذه القيمة نستطيع حساب فترات الثقة للوسط الحسابى لنسبة محتوى الكليسيوم فى المجتمع .

## الاختبارات الإحصائية :

يهيمن هنا إجراء الاختبارين الآتين :

( أولا ) اختبار ما إذا كان محتوى الكلسيوم يختلف من نبات إلى آخر .

وهذا يعنى اختبار الفرض الصفرى  $H_0 = 0$  - راجع البند ( ٨ - ١٤ ) .  
ومن الجدول ( ١٥ - ٧ ) نرى أنه إذا كان هذا الفرض صحيحا فإن  $E_1$  ،  $E_2$  ،  
يكونان تقديرين مستقلين لنفس التباين وإذن الاختبار المناسب لهذا الفرض هو  
اختبار ف بالصورة الآتية :

$$F = \frac{E_1}{E_2} \quad \text{بدرجتى حرية ك - ١ ، ا ك - ك} \quad (٣٤)$$

$$\therefore \text{ فى } F = \frac{2,5201}{0,3288} = 7,66^{**} \quad \text{بدرجتى حرية ٣ ، ٨}$$

وهذه القيمة تزيد عن القيمة الحرجة  $F_{[8,3],0.01} = 7,59$  مما يدعونا إلى رفض  
الفرض الصفرى عند المستوى ٠,٠١ واستنتاج أن محتوى الكلسيوم لا يتساوى  
فى جميع النباتات .

( ثانيا ) اختبار ما إذا كان محتوى الكلسيوم يختلف من ورقة إلى أخرى . وهذا  
يعنى اختبار الفرض الصفرى  $H_0 = 0$  ومن الجدول ( ١٥ - ٧ ) نرى أنه إذا  
كان هذا الفرض صحيحا فإن  $E_1$  ،  $E_2$  ،  $E_3$  يكونان تقديرين مستقلين للتباين  $H_0$   
وإذن الاختبار المناسب هو :

$$F = \frac{E_1}{E_2} \quad \text{بدرجتى حرية ا ك - ك ، ب - ا ك} \quad (٣٥)$$

بلد رجتى حرية ٨ ، ١٢

$$\therefore F = \frac{0.3288}{0.0067} = 49$$

وهذه القيمة تزيد كثيرا عن القيمة الحرجة  $F_{[12, 8]} = 4.50$  مما يجعلنا نرفض الفرض الصفرى عند مستوى عالى من الدلالة ونحكم بأن محتوى الكلسيوم لا يتساوى فى جميع الأوراق .

## SYSTEMATIC SAMPLE

## (١٥ - ٦) العينة المنتظمة

تتخذ المعاينة المنتظمة الأسلوب المبين بالمثل الآتى . نفرض أن أحد الجيولوجيين مهم بدراسة محتوى المعادن الثقيلة فى البطانة المعدنية mineral suit لمنطقة رملية بها  $h = 1000$  طبقة بارزة من الصخور outcrops ممتدة فى صف على سطح الأرض ، والمطلوب اختيار عينة حجمها  $n = 50$  صخرة لتحليلها وفصل المعادن منها ، على أن تؤخذ وحداتها بحيث تكون موزعة توزيعا متعادلا على المنطقة . لتحقيق ذلك نرقم الصخور من ١ إلى ١٠٠٠ ثم نقسمها إلى  $n = 50$  قسما بحيث يحتوى كل قسم على نفس العدد من الصخور أى على  $k = 20$  صخرة ( $1000 \div 50 = 20$ ) . نختار عشوائيا عددا يقع بين ١ ، ٢٠ ليحدد رقم الصخرة التى تؤخذ من القسم الأول . فمثلا إذا كان العدد العشوائى ٣ يكون هذا العدد هو رقم الصخرة الأولى فى العينة ، ثم نأخذ بانتظام أعدادا يزيد كل منها عن سابقه بمقدار ٢٠ فتكون أرقام الصخور التى تدخل فى العينة هى ٣ ، ٢٣ ، ٤٣ ، ٦٣ ، ٨٣ ، ١٠٣ ، ١٢٣ ، ١٤٣ ، ١٦٣ ، ١٨٣ . أما إذا كان العدد العشوائى الذى اختير فى البداية هو ١٥ فإن أرقام الصخور تكون فى هذه الحال ١٥ ، ٣٥ ، ٥٥ ، ٧٥ ، ٩٥ ، ١١٥ ، ١٣٥ ، ١٥٥ ، ١٧٥ ، ١٩٥ .

فى هذا المثال كان حجم المجتمع  $N$  مضاعفا صحيحا لعدد الأقسام ( $N = nh$ ) ولذلك فإن أى عينة تختار بهذه الطريقة تأخذ نفس الحجم  $n$  . أما إذا

كان  $h \neq u$  ك فإن العينات لا تكون جميعها من حجم واحد بل قد يزيد حجم بعضها بواحد عن البعض الآخر . فمثلا نفرض أن  $h = 23$  ،  $u = 5$  .  
إن العينات الخمسة التى يمكن اختيارها تكون أرقام وحداتها كما فى الجدول (١٥ - ٨) الآتى :

الجدول (١٥ - ٨)

(١)	(٢)	(٣)	(٤)	(٥)
١	٢	٣	٤	٥
٦	٧	٨	٩	١٠
١١	١٢	١٣	١٤	١٥
١٦	١٧	١٨	١٩	٢٠
٢١	٢٢	٢٣		

حيث نلاحظ أن حجم كل من العينات الثلاث الأولى  $u = 5$  بينما حجم كل من العيتين الأخيرتين  $h = 23$  . وهذه الحقيقة تسبب إزعاجاً فى تحليل بيانات العينة المنتظمة .

إذا كان  $h = u$  ك (حيث تتساوى أحجام العينات التى تؤخذ) يكون الوسط الحسابى للعينة تقديراً غير متحيز للوسط الحسابى للمجتمع . أما إذا كان  $h \neq u$  ك فإن هذا التقدير يكون متحيزاً ، غير أنه يمكن إزالة هذا التحيز بإعطاء احتمالاً أكبر لاختيار بعض العينات . ففى الجدول (١٥ - ٨) إذا أعطينا الاحتمال  $\frac{5}{23}$  لاختيار كل من العينات الثلاث الأولى والاحتمال  $\frac{4}{23}$  لاختيار كل من العيتين الأخيرتين فإن الوسط الحسابى للعينة يكون حيزتد تقديراً غير متحيز لمتوسط المجتمع . ( لاحظ أن  $0.1 = \frac{8+10}{23} = 2 \times \frac{4}{23} + 3 \times \frac{5}{23}$  )

أما تقدير تباين المجتمع أو تقدير الخطأ المعياري لتوزيع المعاينة للوسط الحسابي فلا توجد طريقة موثوق بها لإيجادها من بيانات مشاهدة في عينة وهذا هو العيب الرئيسي في المعاينة المنتظمة .

أما العيب الثاني فإن العينة المنتظمة تكون متحيزة ولا تعبر تعبيرا صادقا عن المجتمع إذا كان هناك نوع من الاختلافات الدورية أو الموسمية في وحدات المجتمع خاصة إذا حدث أن كانت الوحدات المختارة قريبة من مراكز هذه الاختلافات .

ومن الواضح أن العينة العشوائية المنتظمة أسهل وأسرع في اختيارها من أى عينة أخرى وأقل تعرضا للخطأ إذ يكفي تحديد عدد عشوائى واحد . ولذلك فهى تستخدم حين تكون أقل تكلفة بكثير من أى طريقة أخرى للمعاينة ، أو حين يكون المطلوب تغطية المجتمع بشكل متعادل فهى في هذه الحالة تعطى نتائج أكثر دقة من العينة العشوائية البسيطة .

هذا مع ملاحظة أنه في المعاينة المنتظمة يكون لكل وحدة من وحدات المجتمع نفس الفرصة في الدخول في العينة ، وهى في هذه الصفة تشبه المعاينة العشوائية البسيطة . غير أن احتمال الحصول على عينة منتظمة من حجم ما لا يكون مساويا لاحتمال الحصول على عينة منتظمة أخرى من نفس الحجم ( اختبرت بتغير العدد العشوائى الابتدائى ) كما هو الحال في المعاينة العشوائية البسيطة ، وهذا فرق كبير بين هذين النوعين من المعاينة . والواقع أن العينة المنتظمة ليست عشوائية إلا في اختيار العدد الابتدائى مما يعوق عملية التحليل الاحصائى .

#### Area Sampling

#### (١٥ - ٧) المعاينة المساحية

المعاينة المساحية هى تلك التى تختار فيها العينات من مسطح من الأرض ، وهى تستخدم في كثير من الدراسات السكانية وفي علوم الزراعة والجيولوجيا وغيرها .

طريقة المعاينة المساحية ليس بها جديد من حيث المبدأ إذ تؤخذ العينات بنفس الطرق سابقة الذكر بحسب طبيعة الدراسة التي تجرى ، فقد تكون عشوائية بسيطة أو طبقية أو على مراحل أو منتظمة . ولا يحتاج الأمر إلا إلى إدخال تعديل في خطة المعاينة يمكننا من تحديد الوحدات التي تدخل في العينة .

ومن الإجراءات المفيدة هنا رسم خريطة مصغرة للمسطح المعطى على مستوى يحدد بمحورين متعامدين أحدهما يعين مثلاً الشمال والجنوب والآخر يعين الشرق والغرب مع تحديد نقطة أصل مناسبة . وعند المعاينة نختار النقط ( أو القطع أو المربعات ) التي تدخل في العينة عن طريق الاختيار العشوائى لأزواج من الأعداد نتخذ كإحداثيات للنقط التي نحدد على الخريطة ومن ثم على المسطح الأصل . والمتاد اختيار وحدات هذه الأزواج من الأعداد عن طريق جداول الأرقام العشوائية .

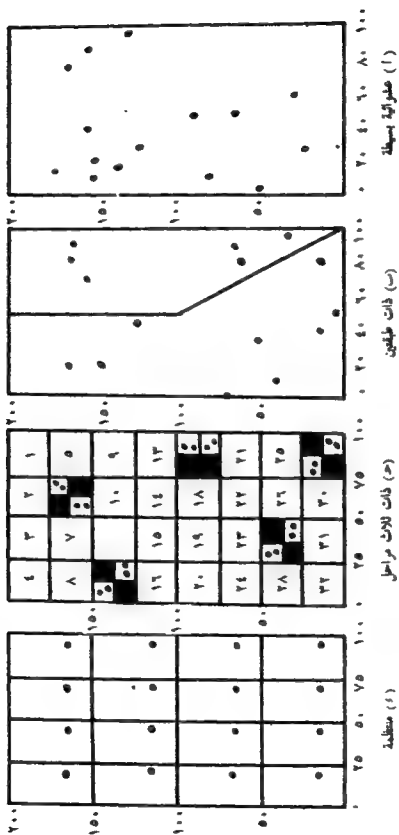
#### مثال ( ١٥ - ٩ ) :

لدينا منطقة من الأرض مستطيلة الشكل عرضها ١٠٠ متراً وطولها ٢٠٠ متراً ونريد دراسة بعض الخواص الكيميائية لتربة هذه الأرض عن طريق اختيار عينة منها ، في الحالات الأربع الآتية :

(أولاً) : اختيار عينة عشوائية بسيطة من ١٥ نقطة .

نستخدم جداول الأرقام العشوائية لاختيار ١٥ عدداً عشوائياً يقع بين ٠ ، ١٠٠ مثلاً ٧٤ ، ١٠ ، ١١ ، ٤٤ ، ١٦ ، ... ثم لاختيار ١٥ عدداً عشوائياً يقع بين ٠ ، ٢٠٠ مثلاً ١٦٨ ، ١٥٠ ، ١٧٤ ، ٩١ ، ١٣٦ ، ... وبهذا يتحدد لنا ١٥ زوجاً من الأعداد هي (٧٤ ، ١٦٨) ، (١٠ ، ١٥٠) ، (١١ ، ١٧٤) ، (٤٤ ، ٩١) ، (١٦ ، ١٣٦) ، ... كل منها يحدد نقطة في المستوى ، فتكون النقطة الناتجة هي وحدات العينة المطلوبة . انظر الشكل ( ١٥ - ١ - أ ) .





الشكل (١٥ - ١) : عينات عشوائية مساحية

(ثانياً) : اختيار عينة طبقية من ١٥ نقطة إذا كان من المعروف أن الأرض مقسمة إلى طبقتين مختلفتين النسبة بينهما ٢ : ٣ على وجه التقريب .

نسحب من الطبقتين عيتين عشوائيتين بسيطتين يتناسب حجمهما مع حجمي الطبقتين ، أى نسحب  $10 \times \frac{2}{5} = 4$  وحدات من الطبقة الأولى ،  $10 \times \frac{3}{5} = 6$  وحدات من الطبقة الثانية . وعلى ذلك نختار ٦ أزواج من الأعداد العشوائية بحيث تقع النقط الممثلة لها في الطبقة الأولى ، ونختار ٩ أزواج من الأعداد العشوائية بحيث تقع النقط الممثلة لها في الطبقة الثانية فنحصل مثلاً على النقط الآتية :

للطبقة الأولى : (٩١ ، ١٦٣) ، (٩٥ ، ٣٥) ، (٨١ ، ٦١) ، (٨٠ ، ١٦٣) ، (٩٠ ، ٦٧) ، (٧٠ ، ١٥٥) .

للطبقة الثانية : (١٩ ، ١٦٦) ، (٤١ ، ١٦) ، (٠ ، ٦٩) ، (٣٢ ، ٥٣) ، (١٩ ، ١٤٧) ، (٥٢ ، ٥) ، (١١ ، ٤٠) ، (٤٥ ، ١٢٧) ، (٨٠ ، ١٥) .

انظر الشكل (١٥ - ١ - ب) .

(ثالثاً) : اختيار عينة عشوائية من ٢٠ نقطة تؤخذ على ثلاث مراحل .

نقسم الأرض إلى عدد من المربعات المتساوية المساحة . فمثلاً إذا أخذنا طول المربع ٢٥ فإن الأرض تنقسم إلى  $4 \times 8 = 32$  مربعا . نرقم هذه المربعات من ١ إلى ٣٢ . نبدأ بأخذ عينة عشوائية من ٥ مربعات ، مثلاً المربعات ذوات الأرقام ١٧ ، ٢٧ ، ٦ ، ٢٩ ، ١٢ وهذه هي المرحلة الأولى (الوحدات الابتدائية) . نقسم كلا من هذه المربعات الخمسة إلى ٤ أجزاء متساوية المساحة ونأخذ من كل مربع جزعين عشوائياً ، وهذه هي المرحلة الثانية وتحتوى على  $5 \times 2 = 10$  أجزاء كل منها ربع مربع . وأخيراً نختار من كل ربع مربع نقطتين عشوائياً فنحصل على ٢٠ نقطة هي التي تمثل العينة المطلوبة - انظر الشكل (١٥ - ١ - ج) .

(رابعا) : اختيار عينة منتظمة من ١٦ نقطة .

نقسم كلا من الطول والعرض إلى ٤ أقسام متساوية الطول فنحصل على ١٦ قسما كل منها مستطيل عرضه ٢٥ مترا وطوله ٥٠ مترا . نحدد عددا عشوائيا بين ٠ ، ٢٥ وليكن ٢٢ ونحدد عددا عشوائيا بين ٠ ، ٥٠ وليكن ١٨ فيكون إحداثيا النقطة الابتدائية (٢٢ ، ١٨) . نحدد النقط الأخرى بانتظام بحيث تبعد كل نقطة عن سابقتها بمسافة قدرها ٢٥ على المحور الأفقى ، ٥٠ على المحور الرأسى . انظر الشكل (١٥ - ١ - ٥) .

#### (١٥ - ٨) العينات غير الاحتمالية :

نعلم أن العينة غير الاحتمالية هى تلك التى نأخذها من المجتمع دون أن نعرف احتمالات دخول وحدات المجتمع فيها ومن ثم لا نستطيع إخضاعها لقواعد الاحتمالات ولا أن نطبق عليها الاختبارات الإحصائية سالفة الذكر . ومع ذلك لا يجوز التقليل من أهمية هذه العينات فكثير منها له استخدامات هامة ويمكن أن تعطى مؤشرات مفيدة عن المجتمعات التى تؤخذ منها .

ومن هذه العينات ما يسمى بالعينة الغرضية Purposeful sample وهى تلك العينة غير العشوائية التى لا تختار بهدف التحليل الإحصائى المتبادل بل لأداء مهمة أو غرض محدد كما هو الحال فى البحوث الاستطلاعية لتقدير تكاليف البحث أو تلمس المشكلات المتوقعة أو لتدريب المساعدين على عملية جمع البيانات . وهنا يختار الباحث الجزء من المجتمع القريب من متناول يده دون تحمل مشقة المعاينة العشوائية .

وهناك ما يسمى بالعينة بالحصص Quota sample وهذا النوع تستخدمه كثير من المؤسسات الصحفية ومعاهد استطلاع الرأى ، ومن أشهرها معهد جالوب Gallup بالولايات المتحدة الأمريكية الذى يستشف نتائج الانتخابات العامة قبل

لإجرائها بسرعة وتكاليف قليلة ، فيطلب من عدد من العاملين استطلاع رأى عدد معين من الناس ( حصة ) في أحد الأحياء أو المناطق فيقوم كل عامل بسؤال من يصادفه من الناس في المكان المحدد له حتى يتم الحصة المنوطة به .

كما أن هناك عينات اضطرارية كما هو الحال في عينة تتألف من متطوعين في الدراسات التي تكون فيها القياسات أو التجارب متعبة أو غير مستحبة أو تحتمل الضرر للأفراد الذين تجرى عليهم الدراسة .

وهناك ما يسمى بعينة التفتيش Search sample وهي تلك التي تهدف إلى التفتيش عن معلومات جديدة كصيد أنواع جديدة من الحشرات أو القواقع أو الصخور المعدنية ، أو الكشف عن رواسب جيرية تصلح لصناعة الأسمت ، أو التنقيب عن الآبار والمياه الجوفية مما يفتح آفاقا جديدة للدراسة النظرية والتطبيقات العملية .

## ملحق (١) أجوبة التمارين

تمارين (١) :

$$(٢) \quad ٦٥,٠ \quad ٠,٠٦٤ \quad ١٤٤,٠ \quad ٣,٠ \quad ٣,٤٥ \quad ٢٥,٠$$

$$(٣) \quad ٤٩,٤٨ \quad ٤٩,٥$$

$$(٤) \quad \frac{1}{3} \quad \frac{2}{10} \quad \frac{3}{10}$$

$$(٥) \quad \frac{19}{20}$$

تمارين (٢ - ١) :

$$(١) \quad ١٠,٩ \quad ١٣,١٥٧ \quad ١٢,٠٧$$

$$(٢) \quad ٤٨,٨٩ \quad ١٠,٣٢ \quad ٢١,١١$$

$$(٣) \quad ٣,٠٢٥ \quad ١,٧٦٥$$

$$(٤) \quad ٤ \text{ م م } ١٠^{-1} \quad , \quad ٠,٣٤٦ \text{ م م } ١٠^{-1} \quad , \quad ٦٢,٦\% \quad , \quad ٤$$

$$(٦) \quad ٢,٢٧٥ = \text{ر} \quad , \quad ٢,٥ = \text{ر} \quad , \quad ٢,٩٨ = \text{ر}$$

تمارين (٢ - ٢) :

(١) التوزيع ملتوى إلى اليمين .

(٢) في توزيع غير المدخنين يتجمع عدد كبير في وسط التوزيع ( بين ٢١,١٩ )

ويتناقص هذا العدد تدريجياً عند الطرفين ، وبالعكس في توزيع المدخنين .

(٣) نعم الفيران تتعلم من التدريب — في الفيران المدربة يتجمع عدد كبير من

الفروق حول القيمة ٤ وهي في وسط التوزيع ويتناقص العدد تدريجياً في الطرفين ،

وبالعكس في الفيران غير المدربة .

تقارين (٣ - ١) :

- (١)  $9,77 \times 10^{-14}$   
 (٢)  $0,3110$  ،  $\bar{c} =$   
 (٣)  $0,672$   
 (٤) وفق توزيع ذى الحدين دليلا ٣ ، ٠,٢ ثم قارن التكرارات المتوقعة بالتكرارات المشاهدة .

تقارين (٣ - ٢) :

- (١)  $0,4868$  ،  $0,3000$  ،  $0,1262$  ،  $0,0360$   
 (٢)  $0,9048$   
 (٣)  $0,0767$   
 (٤)  $\bar{c} = 0,4363$  ،  $c^2 = 0,9709$   
 ٣٨,٧ ١٦٦,١ ٣٦,٢ ٥,٣٠ ٠,٦ ٠,١ ٠,٠ ٠,٠ ٠,٠  
 (٥)  $\bar{c} = 0,4643$  ،  $c^2 = 0,269$   
 التكرارات المتوقعة ٧٠,٤ ٣٢,٧ ٧,٦ ١,٢ ٠,١  
 (٦)  $0,3679$  ،  $0,2642$

تقارين (٤) :

- (١) - أ -  $0,4332$  ،  $0,0490$  ،  $0,2296$  ،  $0,8742$   
 - ب -  $0,1004$  ،  $0,0080$   
 (٢)  $0,9820$  ،  $0,6826$   
 (٣)  $0,9999$  ،  $0,0436$

تقارين (٥) :

- (١) - أ -  $2,179$  ،  $2,423$  ،  $2,921$  ،  $1,960$   
 - ب -  $31,410$  ،  $07,34$  ،  $47,212$  ،  $3,841$   
 - ج -  $3,08$  ،  $4,10$  ،  $4,17$  ،  $0,70$

٠,٠٢٥	٠,٠٥	(٢) - أ -
٠,٥	٠,٠١	- ب -
٠,٠٢٥	٠,٠١	- ج -

تمارين (٦ - ١) :

(١) ت  $0,05 \times$  نرفض الفرض الصفري في الحالتين - (٠,٦٨٥٤ ، ٠,٩١٦٦)

(٢) ت = ٥,٦٩٣ نرفض الفرض الصفري - (٨٨,٠٧٢ ، ٨٢,٥٣)

(٣) (١٠,٨١ ، ٧,٩٩)

(٤) ع = ٠,٦٩٢٨ ، م = ٠,٤٣٧٩ ، ت = ٢٧٤ ، الفرق ليس ذا دلالة

(٥) ع = ١٧٢,٣٥٧ ، م = ٧,٠٥٨ ، ت = ١,٠١٢ ، الفرق ليس ذا

دلالة

(٦) نرفض الفرض الصفري في الحالتين - (٤٥,٧٨ ، ٤٢,٨٩) -

(٤٦,٢٥ ، ٤٢,٤١)

تمارين (٦ - ٢) :

(١) لا يوجد دليل للشك في نظرية مندل  $\chi^2 = ٠,١٣٧$  ،  $\chi^2 = ٠,٤٥٠$

(٢)  $\chi^2 = ١٠,٥٢٦$  الفرق بين المصائد ذو دلالة عند المستوى ٠,٠٥

(٣)  $\chi^2 = ١,٩٤$  عدد المواليد غير ثابت خلال شهور السنة (لا يعتمد

الاختبار لأن هناك غلط ) .

(٤)  $\chi^2 = ٤,٣٢$  نقبل الفرض الصفري عند ٠,٠١ ونرفضه عند ٠,٠٥

المجتمع ربما يفضل النوع أ

(٥)  $\chi^2 = ٥,٢٢٤$  نقبل الفرض الصفري أن الزهر غير متحيز عند المستوى

٠,٠٥

(٧)  $\chi^2 = ٢٣,٦٨٣$  حيوية الحبوب غير مستقلة عن المعالجة الحرارية .

(٨)  $\chi^2 = ٢,٣٨$  اللواء ليس له تأثير بناء على هذه التجربة .

$$(9) (111,74, 4,80)$$

$$(10) (1700,80, 683,91)$$

$$(11) 9,87 = \chi^2$$

تمارين (٦ - ٣) :

$$(1) (0,770, 0,316) (3) (0,233, 0,087) (2) (0,386, 0,294)$$

$$(4) ص. = - 6,284 \text{ نرفض } (5) ص. = 2,948 \text{ نرفض } ف$$

تمارين (٦ - ٤) :

$$(1) 68 = n (2) (33,00, 30,90) , 96 = n (3) 545 = n$$

تمارين (٦ - ٥) :

حدا المراقبة ٧,٣

تمارين (٧) :

$$(1) 0,998 = 0, 42,08 = 1, 48,92 = 1$$

$$(2) 0,94 = 0, 20,008 = 1 : أولا$$

$$\text{ثانيا : } 0,88 = 0, 22,138 = 1, 20,862 = 1$$

$$(3) 0,12 = 0 \text{ تقريبا .}$$

تمارين (٨ - ١)

مصدر التباين	د.ح.	قليل التباين	ف
(١) بين الأقسام	٢	٣٣,٠٨٥	$1 >$
داخل الأقسام	٩	٥٦,٥٣	
الكل	١١		
$\mu_1 = \mu_2 = \mu_3$			
(٢) بين الأقسام	٣	٦٠٢,٥٧	$2,21$



داخل الأقسام	٣٧٧٨	٣٣	١١٤,٤٨
الكلى	٥٥٨٥,٧٣	٣٦	
(٣) بين الأقسام	٠,٥٠٥	٢	٠,٢٥٢٥ > ١
داخل الأقسام	٤,٧٠٠	٩	٠,٥٢٢٢
الكلى	٥,٢٠٥	١١	

(٤) ف = ٥,٩٣ هناك دليل على وجود فروق بين المعالجات .

(٥) ف = ٣,٨٩٩ نرفض ف٠ ، تختلف أنواع الخرسانة في متوسط امتصاصها للرطوبة .

(٦) (أولاً) ف = ٢٣,٢٧٥ نرفض القول بتساوى أطوال الدورات الثلاث .  
(ثانياً) ف = ٣,٦٧٦ نقبل الفرض الصفري عند ٠,٠٥ ونرفضه عند ٠,٠١ .

(٧)  $\chi^2 = ٠,٠٣٩٣$  (أ) هناك اختلاف جوهري بين مجموعتي العلاج ومجموعة المراقبة .

(ب) ليس هناك تلاف جوهري بين نوعي العلاج .

#### تقارين (٨ - ٢)

(١) أولاً : ٣,٦٧ ، ٤,٢٢ كل من العاملين ذو دلالة عالية .

ثانياً : ٤,٢٦ نرفض عند ٠,٠٥ .

(٢) للغذاء ف = ٦,٦٢ متوسطات الكلسترول ليست متساوية في أنواع الغذاء .

للمعامل ف = ٤,٨٦ متوسطات المعامل ليست متساوية .

(٣) عامل الأيام ف = ٢,٣٠ ليست ذات دلالة .

عامل العمق ف = ٢٨٥١,١ ذو دلالة عالية . درجة الحرارة تنخفض بزيادة العمق .

### تمارين (٨ - ٣)

١ >	٢,٧٢٧	١	٢,٧٢٧	بين الأعمدة (السلالات)
٤,٦٩٩٩	٤٩,٣٨١	٢	٩٨,٧٦١	بين الصفوف (الملوحة)
١,١٥٢٣	١٢,١٠٨	٢	٢٤,٢١٥	تفاعل
	١٠,٥٠٧	١٨	١٨٩,١٢٤	خطأ
		٢٣	٣١٤,٨٢٧	كل

### تمارين (٨ - ٤)

٨,٥٢٩	٥٨٩,٦٣	١	٥٨٩,٦٣	(١) بين القسمين (الأعمدة)
,٤٦٤	٣٢,٠٦	١٤	٤٤٨,٨٧	بين الأفراد (الصفوف)
	٦٩,١٣	١٤	٩٦٧,٨٧	الخطأ
		٢٩	٢٠٠٦,٣٧	الكل

(٢) ميزان الطبيب يعطى قراءات أعلى - (٢,٧٥٧٢ ، ٠,٠٤٢٨) .

### تمارين (٨ - ٥)

٠,٩	٥,٧٥	٣	١٧,٢٥	(١) بين الصفوف
٥,٩	٣٨,٢٥	٣	١١٤,٧٥	بين الأعمدة
٩,٠	٥٨,٢٥	٣	١٧٤,٧٥	بين المعالجات
		٦	٣٩,٠٠	الخطأ
		١٥	٣٤٥,٧٥	الكل

تقارین (۸ - ۶)

(۱) ثانيا :

مصدر التباين	۲۲	د ح	ط ۲	ف
بين الأقسام	۳۳۱۲	۴	۸۲۸	۲۰,۷
↓	۴۰,۵	۱	۴۰,۵	۱۰,۱۲۵
↓	۲۸۸۰	۱	۲۸۸۰	۷۲
↓	۲۲,۵	۱	۲۲,۵	۱>
↓	۴,۵	۱	۴,۵	۱>
الخطأ	۱۸۰۰	۴۵	۴۰	
الکلی	۵۱۱۲	۴۹		

ثالثا : لأى مقارنة  $\chi^2 = ۸ -$  القيمة الحرجة ۱۱,۰۷۱

(۲) - أولا

مصدر التباين	۲۲	د ح	تقدير التباين	ف
بين الأعمدة	۳۱۱۴۱,۱۹	۳	۱۰۳۸۰,۳۹۶	۳۳,۲۶۴
بين الصفوف	۴۶۰,۰۶	۲	۲۳۰,۰۳	۱>
تفاعل	۳۴۹۹,۷۲	۶	۵۸۳,۲۸۷	۱,۸۶۹
الخطأ	۷۴۸۹,۳۴	۲۴	۳۱۲,۰۵۶	
الکلی	۴۲۵۹۰,۳۱	۳۵		

ثانيا : متوسطات الأعمدة ٩٤,٣٣٣ ، ٩٧,٦٦٧ ، ١٤٥,٨٨٩ ، ١٦١,٥٥٦  
 ثالثا : للمقارنات البعدية للأعمدة ، القيمة الحرجة ٣١,٣٣٦

#### تمارين (٩ - ١)

(١)  $\hat{u} = ١٢,٠٧ - ٧,٩٩ \cdot \hat{x}$  ،  $\hat{u} = ٢,٣٥٨$  ،  $\hat{u} = ٢,٧٨٥$  ،  
 (٣٤,٢٣ ، ٢٢,١٢)

(٢)  $\hat{u} = ١,٩٣ + ١٨٤,٦٦ \cdot \hat{x}$  ،  $\hat{u} = ٥٩,٧٤١$  ،  $\hat{u} = ٠,١٠٢$  ،  
 هناك علاقة خطية .

(٤)  $\hat{u} = ٥,٢٨ + ٠,٠٠١٩ \cdot \hat{x}$  ،  $\hat{u} = ٠,٩٩٥$  ،  $\hat{u} = ٠,٩٤٥$  ،  
 لا توجد علاقة خطية .

(٦)  $\hat{u} = ٠,١٤٧$  لا توجد علاقة خطية .

#### تمارين (٩ - ٢)

(١)  $\hat{u} = ٢٧,٨٦ - ٠,٥٩٩٩٦ \cdot \hat{x}$  ،  $\hat{u} = ٢٥٧,٤٤$  الانحراف عن  
 الخطية ذو دلالة .

في  $\hat{u} = ٩٥,٨٣$  نرفض  $\beta = ٠$  . هناك علاقة خطية ولكنها ليست أحسن  
 العلاقات .

(٢) العلاقة الخطية تعبر تعبيرا جيدا عن العلاقة الحقيقية بين المتغيرين .

(٤)  $\hat{u} = ٨,٧٥$   $\hat{u}$  ليست مستقلة عن  $\hat{x}$  .

$\hat{u} = ٩,١٥$  يوجد انحراف عن الخطية .

$\hat{u} = ١٠,٠٠$  هناك انحدار خطي ولكنه ليس أفضل العلاقات .

#### تمارين (١٠)

(١)  $\hat{u} = ٠,٤١٨$  ،  $\hat{u} = ١,٤٥٥$  لا توجد علاقة خطية .

- (٢)  $r = 0.8983$  ،  $t = 11.1177$  توجد علاقة خطية .  
 (٤)  $r = 0.128$  ،  $t = 365$  لا توجد علاقة خطية .  
 (٥)  $r = 0.9787$  ،  $t = 12.6$  توجد علاقة خطية سالبة .  
 (٦)  $r = 0.794$  ، القيمة الحرجة عند  $n = 10$  والمستوى  $0.01$  هي  $0.745$  . هناك ارتباط موجب  
 (٧)  $r = 0.98$  ، هناك ارتباط موجب .  
 (٨) معامل الارتباط متساويان في المجتمع .

#### تأريين (١١ - ١)

في  $36.03$  يوجد انحدار خطي ،  $F = 2.14$  المتوسطات متساوية في المجتمع .

المصدر	ب	ح / ا	ب - ح / ا	د ح	ط م	ف
بين الأقسام	—	—	68,697	٢	34,35	٢,١٤
داخل الأقسام	995,١	578	417,100	٢٦	17,04	
الكلي	1288,700	802,903	485,797	28		

#### تأريين (١١ - ٢)

في  $24.4$  يوجد تأثير خطي ،  $F = 4.79$  المتوسطات ليست متساوية .  
 $n = 0.34$  المتوسطات المعنلة  $23.77$  ،  $27.52$  ،  $25.19$  ،  $24.87$  .

### تمارين (١٢ - ١)

$$\begin{aligned} \hat{S} &= 9,7856 + 0,0017 \cdot S + 0,2174 \cdot T \\ F &= 0,8319 \text{ نرفض } \beta, \beta = 0 \\ T &= 2,10 \text{ نقبل } \beta, 0 = T, 2,87 = \text{نرفض } \beta, 0 = \end{aligned}$$

### تمارين (١٢ - ٢)

$$\begin{aligned} (1) \quad F &= 4,4996 \text{ نرفض } H_0 \\ (2) \quad S_{2-31} &= 0,445, T = 2,98 \text{ نرفض } H_{2-31}, 0 = \end{aligned}$$

### تمارين (١٤ - ١)

$$\begin{aligned} (1) \quad S &= 10, \mu = 16, \sigma = 2,603, \text{ ص} = 2,113 \\ &\text{هناك نمط دوري} \\ (2) \quad \text{الوسيط} &= 30,5, S = 13, \mu = 13, \sigma = 2,396 \\ &\text{ص} = 0,209 \pm \text{لا يوجد دليل ضد الفرض أن العينة عشوائية.} \\ (3) \quad \text{الوسيط} &= 35, S = 10, \mu = 11, \sigma = 2,176 \\ &\text{ص} = 0,23 \text{ لا يوجد دليل ضد الفرض أن العينة عشوائية.} \end{aligned}$$

### تمارين (١٤ - ٢)

$$\begin{aligned} (1) \quad N &= 15, S = 4, L (S \geq 4) = 0,056 \\ &\text{نقبل أن المتوسط يساوي } 0,55 \\ (2) \quad N &= 26, S = 6, \text{ ص} = -22,55 \\ &\text{نرفض الفرض الصفري. العالم الأول أفضل.} \end{aligned}$$

### تمارين (١٤ - ٣)

$$N = 5, S = 1 \text{ الامتناع عن التدخين يزيد الوزن.}$$

#### تمارين (١٤ - ٤)

- (١) ن = ٧ ، ع = ٣  
الأوكسجين ينقص مع العمق
- (٢) ن = ١٠ ، ع = ٧  
وزن القلب يزداد بازدياد ضغط الدم .

#### تمارين (١٤ - ٥)

- (١) ص = ٥,٥٨ نرفض  $\alpha = ٠,١٠$  عند  $\alpha = ٠,١٠$  ويبدو أن مستوى التلوث أكبر في النهر الثاني .
- (٢) ص = ٩,٢٠ نرفض  $\alpha = ٠,١٠$  عند  $\alpha = ٠,١٠$  اللون له تأثير .





ملحق (٢)  
جداول احصائية

المجلد (١)  
٢٥٠٠ من الأرقام العشوائية

	1	2	3	4	5	6	7	8	9	10	
1	48461	14952	72619	73689	52059	37086	60050	86192	67049	64739	1
2	76534	38149	49692	31366	52093	15422	20498	33901	10319	43397	2
3	70437	25861	38504	14752	23757	59660	67844	78815	23758	86814	3
4	59584	03370	42806	11393	71722	93804	09095	07856	55589	46020	4
5	04285	58534	16085	51555	27501	73883	33427	33343	45507	50063	5
6	77340	10412	69189	85171	29082	44785	83638	02583	96483	76553	6
7	59183	62687	91778	80354	23512	97219	65921	02035	59847	91403	7
8	91800	04281	39979	03927	82564	28777	59049	97532	54540	79472	8
9	12066	24817	81099	48940	69554	55925	48379	12866	51232	21580	9
10	69907	91751	53512	23748	65906	91385	84983	27915	48491	91068	10
11	80467	04873	54053	25955	48518	13815	37707	68687	15570	08890	11
12	78057	67835	28302	45048	56761	97725	58438	91528	24645	18544	12
13	05648	39387	78191	88415	60269	94880	58812	42931	71898	61534	13
14	22304	39246	01350	99431	61862	78688	30339	60222	74052	25740	14
15	61346	50269	67005	40442	33100	16742	61640	21046	31909	72641	15
16	66793	37696	27965	30459	91011	51426	31006	77468	61029	57108	16
17	86411	48809	36698	42453	83061	43769	39948	87031	30767	13953	17
18	62098	12825	81744	28882	27369	88183	65846	92545	05065	22655	18
19	68775	06261	54265	16203	23340	84750	16317	88686	86842	00879	19
20	52679	19595	13687	74872	89181	01939	18447	10787	76246	80072	20
21	84096	87152	20719	25215	04349	54434	72344	93008	83282	31670	21
22	63964	55937	21417	49944	38356	98404	14850	17994	17161	98981	22
23	31191	75131	72386	11689	95727	05414	88727	45583	22568	77700	23
24	30545	68523	29850	67833	05622	89975	79042	27142	99257	37349	24
25	52573	91001	52315	26430	54175	30122	31796	98842	37600	26025	25
26	16586	81862	01076	99414	31574	94719	34656	80018	86988	79234	26
27	81841	88481	61191	25013	30272	23388	22463	65774	10029	58376	27
28	43563	66829	72838	08074	57080	15446	11034	98143	74989	26885	28
29	19945	84193	57581	77252	85604	45412	43556	27518	90572	00565	29
30	79374	23796	16919	99691	80276	32818	62953	78831	54395	30705	30
31	48503	26615	43980	09810	38789	66679	73799	48418	12647	40044	31
32	32049	65541	37937	41105	70106	89706	40829	40789	59547	00783	32
33	18547	71562	95493	34112	76895	46766	96395	31718	48302	45893	33
34	03180	96742	61486	43305	34183	99605	67803	13471	09243	29557	34
35	94822	24738	67749	83748	59799	25210	31093	62925	72061	69991	35
36	34330	60599	85828	19152	68499	27977	35611	96240	62747	89529	36
37	43770	81537	59527	95674	76692	86420	69930	10020	72881	12532	37
38	56908	77192	50623	41215	14311	42834	80651	93750	59957	31211	38
39	32787	07189	80539	75927	75475	73965	11796	72140	48944	74156	39
40	52441	78392	11733	57703	29133	71164	55355	31006	25526	55790	40
41	22377	54723	18227	28449	04570	18882	00023	67101	06895	08915	41
42	18376	73460	88841	39602	34049	20589	05701	08249	74213	25220	42
43	53201	28610	87957	21497	64729	64983	71551	99016	87903	63875	43
44	34919	78901	59710	27396	07593	05665	11964	44134	00273	76358	44
45	33617	92159	21971	16901	57383	34262	41744	60891	57624	06962	45
46	70010	40964	98780	72418	52571	18415	64362	90636	38034	04909	46
47	19282	68447	35665	31530	59832	49181	21914	65742	89815	39231	47
48	91429	73328	13266	54898	68795	40948	80808	63887	89339	47938	48
49	97637	78393	33021	05867	86520	45363	43066	09088	64040	09803	49
50	95150	07625	05255	83254	93943	52325	93230	62668	79529	65964	50

المجلد (٧)  
معلومات في الحدين : قس

k	(1) k	(2) k	(3) k	(4) k	(5) k	(6) k	(7) k	(8) k	(9) k	(10) k	(11) k	(12) k	(13) k
0	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	2	3	4	5	6	7	8	9	10	11	12	13
2		1	3	6	10	15	21	28	36	45	55	66	78
3			1	4	10	20	35	56	84	120	165	220	286
4				1	5	15	35	70	126	210	330	495	715
5					1	6	21	56	126	252	462	792	1287
6						1	7	28	84	210	462	924	1716
7							1	8	36	120	330	792	1716
8								1	9	45	165	495	1287
9									1	10	55	220	715
10										1	11	66	286
11											1	12	78
12												1	13
13													1

k	(14) k	(15) k	(16) k	(17) k	(18) k	(19) k	(20) k
0	1	1	1	1	1	1	1
1	14	15	16	17	18	19	20
2	91	105	120	136	153	171	190
3	364	455	560	680	816	969	1140
4	1001	1365	1820	2380	3060	3876	4845
5	2002	3003	4368	6188	8568	11628	15504
6	3003	5005	8008	12376	18564	27132	38760
7	3432	6435	11440	19448	31824	50388	77520
8	3003	6435	12870	24310	43758	75582	125970
9	2002	5005	11440	24310	48620	92378	167960
10	1001	3003	8008	19448	43758	92378	184756
11	364	1365	4368	12376	31824	75582	167960
12	91	455	1820	6188	18564	50388	125970
13	14	105	560	2380	8568	27132	77520
14	1	15	120	680	3060	11628	38760
15		1	16	136	816	3876	15504
16			1	17	153	969	4845
17				1	18	171	1140
18					1	19	190
19						1	20
20							1

تابع الجدول (٣)  
الاحتمالات في توزيع ذي الحدين : د ( س )

n	x	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
9	0	0.530	0.387	0.154	0.040	0.010	0.002					
	1	0.390	0.267	0.102	0.026	0.006	0.001	0.004				
	2	0.083	0.172	0.308	0.367	0.161	0.070	0.021	0.004			
	3	0.006	0.046	0.170	0.367	0.361	0.164	0.074	0.021	0.003		
	4	0.001	0.007	0.006	0.172	0.361	0.366	0.167	0.074	0.017	0.001	
	5		0.001	0.017	0.074	0.167	0.366	0.251	0.172	0.066	0.007	0.001
	6			0.003	0.001	0.074	0.164	0.251	0.207	0.176	0.045	0.000
	7				0.004	0.021	0.070	0.161	0.267	0.302	0.172	0.000
	8					0.004	0.018	0.090	0.186	0.302	0.267	0.000
	9						0.002	0.010	0.040	0.134	0.267	0.000
10	0	0.390	0.260	0.107	0.028	0.006	0.001					
	1	0.261	0.267	0.208	0.121	0.040	0.010	0.002				
	2	0.075	0.194	0.302	0.283	0.131	0.044	0.011	0.001			
	3	0.010	0.067	0.201	0.267	0.215	0.117	0.042	0.009	0.001		
	4	0.001	0.011	0.086	0.200	0.261	0.206	0.111	0.027	0.004		
	5		0.001	0.006	0.108	0.201	0.244	0.201	0.108	0.009	0.001	
	6			0.006	0.027	0.111	0.206	0.251	0.200	0.088	0.011	0.001
	7			0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.067	0.010
	8				0.001	0.011	0.044	0.131	0.223	0.200	0.194	0.075
	9					0.002	0.010	0.040	0.121	0.206	0.267	0.215
	10						0.001	0.004	0.008	0.107	0.260	0.200
11	0	0.390	0.264	0.098	0.020	0.004						
	1	0.260	0.204	0.106	0.023	0.007	0.001	0.001				
	2	0.087	0.212	0.206	0.200	0.080	0.027	0.006	0.001			
	3	0.014	0.071	0.221	0.287	0.177	0.081	0.023	0.004			
	4	0.001	0.010	0.111	0.220	0.226	0.161	0.070	0.017	0.002		
	5		0.002	0.009	0.122	0.221	0.226	0.147	0.067	0.010		
	6			0.010	0.057	0.147	0.226	0.231	0.123	0.009	0.000	
	7			0.002	0.017	0.070	0.161	0.226	0.220	0.111	0.016	0.001
	8				0.004	0.028	0.081	0.177	0.227	0.221	0.071	0.014
	9				0.001	0.006	0.027	0.080	0.200	0.205	0.213	0.067
	10					0.001	0.005	0.027	0.023	0.226	0.204	0.220
	11							0.004	0.009	0.008	0.214	0.200
12	0	0.540	0.382	0.090	0.014	0.002						
	1	0.341	0.277	0.106	0.021	0.007	0.001					
	2	0.090	0.220	0.228	0.108	0.004	0.016	0.002				
	3	0.017	0.085	0.226	0.240	0.142	0.064	0.012	0.001			
	4	0.002	0.021	0.120	0.221	0.212	0.121	0.042	0.008	0.001		
	5		0.004	0.022	0.125	0.227	0.122	0.101	0.029	0.000		
	6			0.014	0.079	0.177	0.226	0.177	0.079	0.016		
	7			0.002	0.020	0.101	0.122	0.227	0.186	0.062	0.004	
	8			0.001	0.006	0.042	0.121	0.212	0.221	0.123	0.081	0.002
	9				0.001	0.012	0.064	0.142	0.240	0.226	0.205	0.017
	10					0.000	0.011	0.064	0.186	0.202	0.220	0.000
	11						0.002	0.017	0.071	0.206	0.277	0.241
	12							0.002	0.014	0.000	0.202	0.240

المجدول (٣)  
الاحتمالات في توزيع ذي الحدين : د ( س )

		p											
n	x	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.96
3	0	0.000	0.010	0.040	0.090	0.200	0.300	0.100	0.000	0.000	0.010	0.003	
	1	0.006	0.180	0.320	0.430	0.480	0.500	0.480	0.430	0.320	0.180	0.096	
	2	0.002	0.010	0.040	0.090	0.180	0.300	0.300	0.400	0.640	0.810	0.900	
3	0	0.007	0.730	0.813	0.363	0.216	0.136	0.064	0.007	0.000	0.001		
	1	0.186	0.943	0.304	0.441	0.433	0.375	0.300	0.180	0.000	0.007	0.007	
	2	0.007	0.007	0.000	0.100	0.300	0.375	0.433	0.441	0.304	0.343	0.136	
3	0	0.001	0.000	0.007	0.004	0.136	0.136	0.316	0.343	0.513	0.730	0.867	
	1	0.015	0.000	0.110	0.340	0.130	0.003	0.000	0.000	0.003	0.000		
	2	0.171	0.303	0.410	0.413	0.346	0.300	0.184	0.070	0.000	0.000	0.004	
3	0	0.014	0.040	0.184	0.364	0.346	0.275	0.346	0.346	0.184	0.040	0.014	
	1	0.004	0.000	0.070	0.184	0.300	0.346	0.413	0.410	0.300	0.171		
	2	0.003	0.003	0.000	0.000	0.003	0.003	0.120	0.340	0.410	0.600	0.810	
3	0	0.774	0.900	0.300	0.166	0.070	0.001	0.010	0.000				
	1	0.304	0.320	0.410	0.300	0.300	0.184	0.077	0.000	0.000			
	2	0.001	0.070	0.304	0.300	0.346	0.313	0.300	0.184	0.001	0.000	0.004	
3	0	0.001	0.000	0.001	0.130	0.300	0.313	0.346	0.300	0.300	0.070	0.001	
	1	0.000	0.000	0.000	0.077	0.184	0.300	0.346	0.410	0.300	0.304	0.304	
	2	0.003	0.010	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.736	0.331	0.300	0.110	0.047	0.016	0.004	0.001				
	1	0.333	0.364	0.300	0.300	0.187	0.004	0.007	0.010	0.000			
	2	0.001	0.000	0.304	0.304	0.311	0.304	0.180	0.000	0.010	0.001		
3	0	0.003	0.016	0.000	0.100	0.270	0.313	0.370	0.180	0.003	0.010	0.000	
	1	0.001	0.010	0.000	0.100	0.304	0.311	0.334	0.346	0.000	0.001		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		
	2	0.000	0.000	0.000	0.000	0.007	0.004	0.107	0.300	0.300	0.344	0.323	
3	0	0.000	0.000	0.000	0.000	0.004	0.016	0.047	0.110	0.300	0.531	0.736	
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	2	0.367	0.373	0.367	0.347	0.131	0.066	0.017	0.004				
3	0	0.011	0.134	0.370	0.310	0.301	0.164	0.077	0.030	0.004			
	1	0.004	0.033	0.110	0.237	0.300	0.373	0.104	0.007	0.000	0.000		

تابع الجدول (٣)  
الاحتمالات في توزيع ذي الحدين : د ( س )

n	x	0.05	0.1	0.3	0.3	0.4	0.5	0.5	0.7	0.5	0.5	0.05
10	0	0.513	0.364	0.066	0.010	0.001						
	1	0.361	0.367	0.179	0.064	0.011	0.002					
	2	0.111	0.365	0.368	0.190	0.046	0.010	0.001				
	3	0.021	0.100	0.366	0.218	0.111	0.036	0.006	0.001			
	4	0.000	0.008	0.104	0.204	0.104	0.027	0.004	0.003			
	5		0.006	0.000	0.180	0.221	0.107	0.006	0.014	0.001		
	6		0.001	0.002	0.100	0.197	0.200	0.121	0.044	0.006		
	7			0.006	0.044	0.121	0.200	0.197	0.103	0.022	0.001	
	8			0.001	0.014	0.006	0.107	0.221	0.100	0.000	0.006	
	9				0.002	0.004	0.007	0.104	0.224	0.104	0.006	0.000
10				0.001	0.000	0.006	0.036	0.111	0.218	0.246	0.100	0.001
11					0.001	0.010	0.045	0.120	0.206	0.245	0.111	
12						0.002	0.011	0.064	0.179	0.267	0.261	
13							0.001	0.010	0.066	0.264	0.215	
14	0	0.480	0.220	0.044	0.007	0.001						
	1	0.260	0.266	0.104	0.041	0.007	0.001					
	2	0.123	0.257	0.220	0.112	0.032	0.006	0.001				
	3	0.026	0.114	0.220	0.194	0.086	0.022	0.002				
	4	0.004	0.025	0.172	0.220	0.155	0.061	0.014	0.001			
	5		0.000	0.000	0.100	0.207	0.122	0.041	0.007			
	6		0.001	0.002	0.120	0.207	0.103	0.002	0.022	0.002		
	7			0.000	0.002	0.157	0.200	0.157	0.002	0.000		
	8			0.002	0.022	0.002	0.103	0.207	0.120	0.022	0.001	
	9				0.007	0.041	0.122	0.207	0.100	0.006	0.006	
10				0.001	0.014	0.061	0.155	0.220	0.172	0.025	0.004	
11					0.006	0.022	0.086	0.194	0.250	0.114	0.006	
12					0.001	0.000	0.032	0.112	0.260	0.267	0.122	
13							0.001	0.007	0.041	0.154	0.236	0.200
14								0.001	0.007	0.044	0.220	0.480
15	0	0.211	0.206	0.026	0.005							
	1	0.200	0.262	0.122	0.031	0.006						
	2	0.126	0.267	0.221	0.082	0.022	0.006					
	3	0.021	0.120	0.250	0.170	0.082	0.014	0.002				
	4	0.000	0.042	0.100	0.210	0.127	0.042	0.007	0.001			
	5	0.001	0.010	0.102	0.206	0.100	0.092	0.024	0.003			
	6		0.002	0.042	0.147	0.207	0.103	0.061	0.012	0.001		
	7			0.014	0.061	0.177	0.100	0.110	0.025	0.002		
	8			0.002	0.026	0.110	0.100	0.177	0.061	0.014		
	9			0.001	0.012	0.061	0.152	0.207	0.147	0.042	0.000	
10					0.000	0.004	0.092	0.100	0.200	0.103	0.010	0.001
11					0.001	0.007	0.042	0.127	0.210	0.106	0.042	0.006
12						0.002	0.014	0.060	0.170	0.250	0.120	0.001
13							0.000	0.022	0.092	0.221	0.267	0.126
14								0.006	0.031	0.122	0.245	0.266
15									0.006	0.005	0.200	0.480

المجدول ( ٤ )  
قيم حتمية

$\lambda$	1	2	3	4	5	6	7	8	9	10
$\sigma^2$	0.36788	0.13534	0.04979	0.01832	0.006738	0.002479	0.000912	0.000335	0.000123	0.000045

$\lambda$	0	1	2	3	4	5	6	7	8	9
0.0	1.0000	0.9900	0.9802	0.9704	0.9608	0.9512	0.9418	0.9324	0.9231	0.9139
0.1	0.9048	0.8958	0.8869	0.8781	0.8694	0.8607	0.8521	0.8437	0.8353	0.8270
0.2	0.8187	0.8106	0.8025	0.7945	0.7866	0.7788	0.7711	0.7634	0.7558	0.7483
0.3	0.7408	0.7334	0.7261	0.7189	0.7118	0.7047	0.6977	0.6907	0.6839	0.6771
0.4	0.6703	0.6636	0.6570	0.6505	0.6440	0.6376	0.6313	0.6250	0.6188	0.6126
0.5	0.6065	0.6005	0.5945	0.5886	0.5827	0.5770	0.5712	0.5655	0.5599	0.5543
0.6	0.5488	0.5434	0.5379	0.5326	0.5273	0.5220	0.5169	0.5117	0.5066	0.5016
0.7	0.4966	0.4916	0.4868	0.4819	0.4771	0.4724	0.4677	0.4630	0.4584	0.4538
0.8	0.4493	0.4449	0.4404	0.4360	0.4317	0.4274	0.4232	0.4190	0.4148	0.4107
0.9	0.4066	0.4025	0.3983	0.3946	0.3906	0.3867	0.3829	0.3791	0.3753	0.3716

$$V_{11} \times V_{12} = V_{12} \times V_{11} \quad : \text{مثال}$$

الجدول (5)

الاحتمالات والاحتمالات المتكاملة في توزيع بواسون

$x$	$\mu = 0.1$		$\mu = 0.2$		$\mu = 0.3$		$\mu = 0.4$		$\mu = 0.5$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
0	0.9048	0.9048	0.8187	0.8187	0.7408	0.7408	0.6703	0.6703	0.6065	0.6065
1	0.0950	0.9953	0.1637	0.9825	0.2222	0.9631	0.2681	0.9384	0.3033	0.9098
2	0.0045	0.9998	0.0164	0.9989	0.0333	0.9964	0.0536	0.9921	0.0758	0.9856
3	0.0002	1.0000	0.0011	0.9999	0.0033	0.9997	0.0072	0.9992	0.0126	0.9982
4	0.0000	1.0000	0.0001	1.0000	0.0003	1.0000	0.0007	0.9999	0.0016	0.9998
5							0.0001	1.0000	0.0002	1.0000

$x$	$\mu = 0.6$		$\mu = 0.7$		$\mu = 0.8$		$\mu = 0.9$		$\mu = 1$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
0	0.5488	0.5488	0.4966	0.4966	0.4493	0.4493	0.4066	0.4066	0.3679	0.3679
1	0.3293	0.8781	0.3476	0.8442	0.3595	0.8088	0.3699	0.7725	0.3679	0.7358
2	0.0988	0.9769	0.1217	0.9659	0.1438	0.9526	0.1647	0.9371	0.1839	0.9197
3	0.0198	0.9966	0.0284	0.9942	0.0383	0.9909	0.0494	0.9865	0.0613	0.9810
4	0.0030	0.9996	0.0050	0.9992	0.0077	0.9996	0.0111	0.9977	0.0153	0.9963
5	0.0004	1.0000	0.0007	0.9999	0.0012	0.9998	0.0020	0.9997	0.0031	0.9994
6			0.0001	1.0000	0.0002	1.0000	0.0003	1.0000	0.0005	0.9999
7									0.0001	1.0000

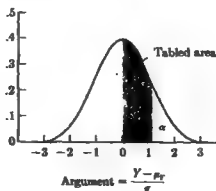
$x$	$\mu = 1.5$		$\mu = 2$		$\mu = 3$		$\mu = 4$		$\mu = 5$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
0	0.2231	0.2231	0.1353	0.1353	0.0498	0.0498	0.0183	0.0183	0.0067	0.0067
1	0.3347	0.5578	0.2707	0.4060	0.1494	0.1991	0.0733	0.0916	0.0337	0.0404
2	0.2510	0.8088	0.2707	0.6767	0.2240	0.4232	0.1465	0.2381	0.0842	0.1247
3	0.1255	0.9344	0.1804	0.8571	0.2240	0.6472	0.1954	0.4335	0.1404	0.2650
4	0.0471	0.9814	0.0902	0.9473	0.1680	0.8153	0.1954	0.6288	0.1755	0.4405
5	0.0141	0.9955	0.0361	0.9834	0.1008	0.9161	0.1563	0.7851	0.1755	0.6160
6	0.0035	0.9991	0.0120	0.9955	0.0504	0.9665	0.1042	0.8893	0.1462	0.7622
7	0.0008	0.9998	0.0034	0.9989	0.0216	0.9881	0.0595	0.9489	0.1044	0.8666
8	0.0001	1.0000	0.0009	0.9998	0.0081	0.9962	0.0298	0.9786	0.0653	0.9319
9			0.0002	1.0000	0.0027	0.9989	0.0132	0.9919	0.0363	0.9682
10					0.0008	0.9997	0.0053	0.9972	0.0181	0.9863
11					0.0002	0.9999	0.0019	0.9991	0.0082	0.9945
12					0.0001	1.0000	0.0006	0.9997	0.0034	0.9988
13							0.0002	0.9999	0.0013	0.9993
14							0.0001	1.0000	0.0005	0.9998
15									0.0002	0.9999
16									0.0000	1.0000



# المجدول (٦)

## المساحات أسفل المنحنى المحلل المعياري

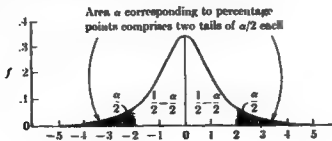
Standard deviation units	0	1	2	3	4	5	6	7	8	9	Standard deviation units
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359	0.0
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753	0.1
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141	0.2
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517	0.3
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879	0.4
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224	0.5
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549	0.6
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852	0.7
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133	0.8
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389	0.9
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621	1.0
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830	1.1
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015	1.2
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177	1.3
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319	1.4
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441	1.5
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545	1.6
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633	1.7
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706	1.8
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767	1.9
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817	2.0
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857	2.1
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890	2.2
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916	2.3
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936	2.4
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952	2.5
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964	2.6
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974	2.7
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981	2.8
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986	2.9
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990	3.0
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993	3.1
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995	3.2
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997	3.3
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998	3.4
3.5	.499767										
3.6	.499841										
3.7	.499892										
3.8	.499928										
3.9	.499952										
4.0	.499968										
4.1	.499979										
4.2	.499987										
4.3	.499991										
4.4	.499993										
4.5	.499997										
4.6	.499998										
4.7	.499999										
4.8	.499999										
4.9	.500000										



Note: The quantity given is the area under the standard normal density function between the mean and the critical point. The area is generally labeled  $\frac{1}{2} - \alpha$  (as shown in the figure). By inverse interpolation one can find the number of standard deviations corresponding to a given area.

المجلد (3)  
العم المخرجة لتوزيع ن

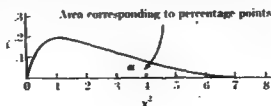
$\alpha$	0.9	0.5	0.4	0.2	0.1	0.05	0.02	0.01	0.001	$\alpha$
1	0.138	1.000	1.376	3.078	6.314	12.706	31.821	63.657	636.619	1
2	0.142	0.816	1.061	1.886	2.920	4.303	6.965	9.925	31.598	2
3	0.147	0.765	1.078	1.633	2.353	3.162	4.941	5.841	12.924	3
4	0.154	0.741	1.071	1.533	2.132	2.716	3.747	4.604	8.610	4
5	0.162	0.727	0.920	1.476	2.015	2.571	3.465	4.032	6.869	5
6	0.161	0.718	0.906	1.440	1.943	2.447	3.143	3.707	5.959	6
7	0.160	0.711	0.896	1.415	1.895	2.365	2.896	3.499	5.406	7
8	0.160	0.706	0.889	1.397	1.860	2.306	2.696	3.355	5.041	8
9	0.159	0.703	0.883	1.383	1.833	2.262	2.621	3.250	4.781	9
10	0.159	0.700	0.879	1.372	1.812	2.228	2.564	3.169	4.587	10
11	0.159	0.697	0.876	1.363	1.796	2.201	2.518	3.106	4.487	11
12	0.158	0.693	0.873	1.356	1.782	2.179	2.461	3.045	4.386	12
13	0.158	0.690	0.870	1.350	1.771	2.160	2.450	3.012	4.271	13
14	0.158	0.687	0.868	1.345	1.761	2.145	2.424	2.977	4.160	14
15	0.158	0.691	0.866	1.341	1.753	2.131	2.402	2.947	4.075	15
16	0.158	0.690	0.865	1.337	1.746	2.120	2.385	2.928	4.015	16
17	0.158	0.689	0.863	1.333	1.740	2.110	2.367	2.908	3.965	17
18	0.157	0.688	0.862	1.330	1.734	2.101	2.352	2.878	3.922	18
19	0.157	0.688	0.861	1.328	1.729	2.093	2.339	2.861	3.883	19
20	0.157	0.687	0.860	1.325	1.725	2.086	2.328	2.845	3.850	20
21	0.157	0.686	0.859	1.323	1.721	2.080	2.318	2.831	3.819	21
22	0.157	0.686	0.858	1.321	1.717	2.074	2.308	2.819	3.792	22
23	0.157	0.685	0.858	1.319	1.714	2.069	2.300	2.807	3.767	23
24	0.157	0.685	0.857	1.318	1.711	2.064	2.292	2.797	3.745	24
25	0.157	0.684	0.856	1.316	1.708	2.060	2.285	2.787	3.725	25
26	0.157	0.684	0.856	1.315	1.706	2.056	2.279	2.779	3.707	26
27	0.157	0.684	0.855	1.314	1.703	2.052	2.273	2.771	3.690	27
28	0.157	0.683	0.855	1.313	1.701	2.048	2.267	2.763	3.674	28
29	0.157	0.683	0.854	1.311	1.699	2.045	2.262	2.756	3.659	29
30	0.157	0.683	0.854	1.310	1.697	2.042	2.257	2.750	3.646	30
40	0.156	0.681	0.851	1.303	1.684	2.021	2.223	2.704	3.551	40
60	0.126	0.679	0.845	1.296	1.671	2.000	2.190	2.680	3.460	60
120	0.126	0.677	0.845	1.289	1.658	1.980	2.158	2.617	3.373	120
$\infty$	0.126	0.674	0.842	1.282	1.645	1.960	2.126	2.516	3.291	$\infty$



(أ) الجدول  
القيم المرحلة لتوزيع

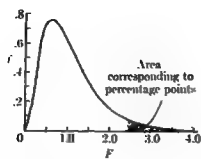
$\alpha$	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	$\alpha$
1	0.000	0.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.482	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15
16	5.142	6.908	9.312	15.338	23.542	26.296	28.845	32.000	34.267	16
17	5.697	7.564	10.085	16.338	24.769	27.587	30.191	33.409	35.718	17
18	6.265	8.231	10.865	17.338	25.989	28.869	31.526	34.805	37.156	18
19	6.844	8.907	11.651	18.338	27.204	30.144	32.852	36.191	38.582	19
20	7.434	9.591	12.443	19.337	28.412	31.410	34.170	37.566	39.997	20
21	8.034	10.283	13.240	20.337	29.615	32.670	35.479	38.932	41.401	21
22	8.643	10.982	14.042	21.337	30.813	33.924	36.781	40.289	42.796	22
23	9.260	11.688	14.848	22.337	32.007	35.172	38.076	41.638	44.181	23
24	9.886	12.401	15.659	23.337	33.196	36.415	39.364	42.980	45.558	24
25	10.520	13.120	16.473	24.337	34.382	37.652	40.646	44.314	46.928	25
26	11.160	13.844	17.292	25.336	35.563	38.885	41.923	45.642	48.290	26
27	11.808	14.573	18.114	26.336	36.741	40.113	43.194	46.963	49.645	27
28	12.461	15.308	18.939	27.336	37.916	41.337	44.461	48.278	50.993	28
29	13.121	16.047	19.768	28.336	39.088	42.557	45.722	49.588	52.336	29
30	13.787	16.791	20.599	29.336	40.256	43.773	46.979	50.892	53.672	30
31	14.458	17.539	21.434	30.336	41.422	44.985	48.232	52.191	55.003	31
32	15.134	18.291	22.271	31.336	42.585	46.194	49.480	53.486	56.329	32
33	15.815	19.047	23.110	32.336	43.745	47.400	50.725	54.776	57.649	33
34	16.501	19.806	23.952	33.336	44.903	48.602	51.966	56.061	58.964	34
35	17.192	20.569	24.797	34.336	46.059	49.802	53.203	57.342	60.275	35
36	17.887	21.336	25.643	35.336	47.212	50.998	54.437	58.619	61.582	36
37	18.586	22.106	26.492	36.335	48.363	52.192	55.668	59.892	62.884	37
38	19.289	22.878	27.343	37.335	49.513	53.384	56.896	61.162	64.182	38
39	19.996	23.654	28.196	38.335	50.660	54.572	58.120	62.428	65.476	39
40	20.707	24.433	29.051	39.335	51.805	55.758	59.342	63.691	66.766	40
41	21.421	25.215	29.907	40.335	52.949	56.942	60.561	64.950	68.053	41
42	22.138	25.999	30.765	41.335	54.090	58.124	61.777	66.206	69.336	42
43	22.859	26.785	31.625	42.335	55.230	59.304	62.990	67.459	70.616	43
44	23.584	27.575	32.487	43.335	56.369	60.481	64.202	68.710	71.893	44
45	24.311	28.366	33.350	44.335	57.505	61.656	65.410	69.957	73.166	45
46	25.042	29.160	34.215	45.335	58.641	62.830	66.617	71.201	74.437	46
47	25.775	29.956	35.081	46.335	59.774	64.001	67.821	72.443	75.704	47
48	26.511	30.755	35.949	47.335	60.907	65.171	69.023	73.683	76.969	48
49	27.249	31.555	36.818	48.335	62.038	66.339	70.222	74.919	78.231	49
50	27.991	32.357	37.689	49.335	63.167	67.505	71.420	76.154	79.490	50

٢٢٩



تابع الجدول (أ)  
القيم المخرجة لتوزيع  $\chi^2$

$\alpha$	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	$\alpha$
51	28.735	33.162	38.560	50.335	64.295	68.669	72.616	77.386	80.747	51
52	29.481	33.968	39.433	51.335	65.422	69.832	73.810	78.616	82.001	52
53	30.230	34.776	40.308	52.335	66.548	70.993	75.002	79.843	83.253	53
54	30.981	35.586	41.183	53.335	67.673	72.153	76.192	81.069	84.502	54
55	31.735	36.398	42.060	54.335	68.796	73.311	77.380	82.292	85.749	55
56	32.490	37.212	42.937	55.335	69.918	74.468	78.567	83.513	86.994	56
57	33.248	38.027	43.816	56.335	71.040	75.624	79.752	84.733	88.237	57
58	34.008	38.844	44.696	57.335	72.160	76.778	80.936	85.950	89.477	58
59	34.770	39.662	45.577	58.335	73.279	77.931	82.117	87.166	90.715	59
60	35.534	40.482	46.459	59.335	74.397	79.082	83.298	88.379	91.952	60
61	36.300	41.303	47.342	60.335	75.514	80.232	84.476	89.591	93.186	61
62	37.068	42.126	48.226	61.335	76.630	81.381	85.654	90.802	94.419	62
63	37.838	42.950	49.111	62.335	77.745	82.529	86.830	92.010	95.649	63
64	38.610	43.776	49.996	63.335	78.860	83.675	88.004	93.217	96.878	64
65	39.383	44.603	50.883	64.335	79.973	84.821	89.177	94.422	98.105	65
66	40.158	45.431	51.770	65.335	81.085	85.965	90.349	95.626	99.331	66
67	40.935	46.261	52.659	66.335	82.197	87.108	91.519	96.828	100.55	67
68	41.713	47.092	53.548	67.334	83.308	88.250	92.689	98.028	101.78	68
69	42.494	47.924	54.438	68.334	84.418	89.391	93.856	99.228	103.00	69
70	43.275	48.758	55.329	69.334	85.527	90.531	95.023	100.43	104.21	70
71	44.058	49.592	56.221	70.334	86.635	91.670	96.189	101.62	105.43	71
72	44.843	50.428	57.113	71.334	87.743	92.808	97.353	102.82	106.65	72
73	45.629	51.265	58.006	72.334	88.850	93.945	98.516	104.01	107.86	73
74	46.417	52.103	58.900	73.334	89.956	95.081	99.678	105.20	109.07	74
75	47.206	52.942	59.795	74.334	91.061	96.217	100.84	106.39	110.29	75
76	47.997	53.782	60.690	75.334	92.166	97.351	102.00	107.58	111.50	76
77	48.788	54.623	61.586	76.334	93.270	98.484	103.16	108.77	112.70	77
78	49.582	55.466	62.483	77.334	94.373	99.617	104.32	109.96	113.91	78
79	50.376	56.309	63.380	78.334	95.476	100.75	105.47	111.14	115.12	79
80	51.172	57.153	64.278	79.334	96.578	101.88	106.63	112.33	116.32	80
81	51.969	57.998	65.176	80.334	97.680	103.01	107.78	113.51	117.52	81
82	52.767	58.845	66.076	81.334	98.780	104.14	108.94	114.69	118.73	82
83	53.567	59.692	66.976	82.334	99.880	105.27	110.09	115.88	119.93	83
84	54.368	60.540	67.876	83.334	100.98	106.39	111.24	117.06	121.13	84
85	55.170	61.389	68.777	84.334	102.08	107.52	112.39	118.24	122.32	85
86	55.973	62.239	69.679	85.334	103.18	108.65	113.54	119.41	123.52	86
87	56.777	63.089	70.581	86.334	104.28	109.77	114.69	120.59	124.72	87
88	57.582	63.941	71.484	87.334	105.37	110.90	115.84	121.77	125.91	88
89	58.389	64.793	72.387	88.334	106.47	112.02	116.99	122.94	127.11	89
90	59.196	65.647	73.291	89.334	107.56	113.15	118.14	124.12	128.30	90
91	60.005	66.501	74.196	90.334	108.66	114.27	119.28	125.29	129.49	91
92	60.815	67.356	75.101	91.334	109.76	115.39	120.43	126.46	130.68	92
93	61.625	68.211	76.006	92.334	110.85	116.51	121.57	127.63	131.87	93
94	62.437	69.068	76.912	93.334	111.94	117.63	122.72	128.80	133.06	94
95	63.250	69.925	77.818	94.334	113.04	118.75	123.86	129.97	134.25	95
96	64.063	70.783	78.725	95.334	114.13	119.87	125.00	131.14	135.43	96
97	64.878	71.642	79.633	96.334	115.22	120.99	126.14	132.31	136.62	97
98	65.694	72.501	80.541	97.334	116.32	122.11	127.28	133.48	137.80	98
99	66.510	73.361	81.449	98.334	117.41	123.23	128.42	134.64	138.99	99
100	67.328	74.222	82.358	99.334	118.50	124.34	129.56	135.81	140.17	100



$\alpha$	1	2	3	4	5	6	7	8	9	10	11	$\alpha$
1	1.61	1.99	2.16	2.29	2.39	2.46	2.51	2.55	2.58	2.61	2.63	.05
.05	6.68	8.00	8.64	9.00	9.22	9.37	9.48	9.57	9.63	9.67	9.70	.025
.01	4.050	5.000	5.400	5.620	5.760	5.860	5.930	5.980	6.020	6.050	6.080	.01
2	1.65	19.0	19.2	19.2	19.3	19.3	19.3	19.4	19.4	19.4	19.4	.05
.05	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	.025
.01	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	.01
3	1.61	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	.05
.05	11.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	.025
.01	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	.01
4	1.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	.05
.05	11.2	10.6	9.98	9.40	9.36	9.20	9.07	8.98	8.90	8.84	8.79	.025
.01	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	.01
5	6.61	5.79	5.41	5.19	5.09	4.99	4.88	4.82	4.77	4.74	4.71	.05
.05	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	.025
.01	18.3	15.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.99	.01
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	.05
.05	8.81	7.36	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	.025
.01	13.7	10.9	9.78	9.15	8.79	8.47	8.26	8.10	7.98	7.87	7.79	.01
7	5.59	4.74	4.35	4.12	3.97	3.87	3.77	3.73	3.68	3.64	3.60	.05
.05	8.07	6.34	5.89	5.52	5.29	5.12	4.99	4.89	4.82	4.76	4.71	.025
.01	12.2	9.35	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	.01
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	.05
.05	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.25	.025
.01	11.3	8.65	7.59	7.01	6.63	6.37	6.16	6.03	5.91	5.81	5.73	.01
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	.05
.05	6.81	5.41	4.72	4.35	4.08	3.92	3.80	3.71	3.63	3.56	3.51	.025
.01	10.6	8.02	6.99	6.42	6.04	5.80	5.61	5.47	5.35	5.26	5.18	.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	.05
.05	6.54	5.14	4.45	4.08	3.81	3.65	3.53	3.43	3.36	3.29	3.24	.025
.01	10.0	7.36	6.35	5.79	5.41	5.19	5.00	4.86	4.74	4.65	4.57	.01

تابع الجدول (٩)  
القيم المرحلة لتوزيع ف

$\alpha$	12	15	20	24	30	40	60	120	$\infty$	$\alpha$
1 .05 .025 .01	244 977 6110	246 985 6160	248 993 6210	249 997 6230	250 1000 6260	251 1010 6290	252 1010 6310	253 1010 6340	254 1020 6370	.05 1 .025 .01
2 .05 .025 .01	19.4 39.4 99.4	19.4 39.4 99.4	19.4 39.4 99.4	19.5 39.5 99.5	19.5 39.5 99.5	19.5 39.5 99.5	19.5 39.5 99.5	19.5 39.5 99.5	19.5 39.5 99.5	.05 2 .025 .01
3 .05 .025 .01	8.74 14.3 27.1	8.70 14.3 26.9	8.66 14.2 26.7	8.64 14.1 26.6	8.62 14.1 26.5	8.59 14.0 26.4	8.57 14.0 26.3	8.55 13.9 26.2	8.53 13.9 26.1	.05 3 .025 .01
4 .05 .025 .01	5.91 8.75 14.4	5.86 8.66 14.2	5.80 8.56 14.0	5.77 8.51 13.9	5.75 8.46 13.8	5.72 8.41 13.7	5.69 8.36 13.7	5.66 8.31 13.6	5.63 8.26 13.5	.05 4 .025 .01
5 .05 .025 .01	4.68 6.52 9.89	4.62 6.43 9.72	4.56 6.33 9.55	4.53 6.28 9.47	4.50 6.23 9.38	4.46 6.18 9.29	4.43 6.12 9.20	4.40 6.07 9.11	4.36 6.02 9.02	.05 5 .025 .01
6 .05 .025 .01	4.00 5.37 7.72	3.94 5.27 7.56	3.87 5.17 7.40	3.84 5.12 7.31	3.81 5.07 7.23	3.77 5.01 7.14	3.74 4.96 7.06	3.70 4.90 6.97	3.67 4.85 6.88	.05 6 .025 .01
7 .05 .025 .01	3.57 4.67 6.47	3.51 4.57 6.31	3.44 4.47 6.16	3.41 4.42 6.07	3.38 4.36 5.99	3.34 4.31 5.91	3.30 4.25 5.82	3.27 4.20 5.74	3.23 4.14 5.65	.05 7 .025 .01
8 .05 .025 .01	3.28 4.20 5.67	3.22 4.10 5.52	3.15 4.00 5.36	3.12 3.95 5.28	3.08 3.89 5.20	3.04 3.84 5.12	3.01 3.78 5.03	2.97 3.73 4.95	2.93 3.67 4.86	.05 8 .025 .01
9 .05 .025 .01	3.07 3.87 5.11	3.01 3.77 4.96	2.94 3.67 4.81	2.88 3.61 4.73	2.84 3.56 4.65	2.81 3.51 4.57	2.77 3.45 4.48	2.75 3.39 4.40	2.71 3.33 4.31	.05 9 .025 .01
10 .05 .025 .01	2.91 3.62 4.71	2.85 3.52 4.56	2.77 3.42 4.41	2.74 3.37 4.33	2.70 3.31 4.25	2.66 3.26 4.17	2.62 3.20 4.08	2.58 3.14 4.00	2.54 3.08 3.91	.05 10 .025 .01

تابع الجدول (٩)  
القيم الخارجة لتعديلي

$\alpha$	1	2	3	4	5	6	7	8	9	10	11	$\alpha$
11	4.84 6.72 9.65 .01	3.98 5.26 7.21 .01	3.59 4.53 6.22 .01	3.36 4.28 5.67 .01	3.20 4.04 5.32 .01	3.09 3.88 5.07 .01	3.01 3.76 4.89 .01	2.95 3.66 4.74 .01	2.90 3.59 4.63 .01	2.85 3.53 4.54 .01	2.82 3.48 4.46 .01	.05 .025 .01 11
12	4.75 6.55 9.33 .01	3.89 5.10 6.93 .01	3.49 4.47 5.95 .01	3.26 4.12 5.41 .01	3.11 3.89 5.06 .01	3.00 3.73 4.82 .01	2.91 3.61 4.64 .01	2.85 3.51 4.50 .01	2.80 3.44 4.59 .01	2.75 3.37 4.30 .01	2.72 3.32 4.22 .01	.05 .025 .01 12
15	4.54 6.20 8.68 .01	3.68 4.77 6.36 .01	3.29 4.15 5.42 .01	3.06 3.80 4.89 .01	2.90 3.58 4.56 .01	2.79 3.41 4.37 .01	2.71 3.29 4.14 .01	2.64 3.20 4.00 .01	2.59 3.12 3.89 .01	2.54 3.06 3.80 .01	2.51 3.01 3.73 .01	.05 .025 .01 15
20	4.35 5.87 8.10 .01	3.49 4.66 5.85 .01	3.10 3.86 4.94 .01	2.87 3.51 4.43 .01	2.71 3.29 4.10 .01	2.60 3.13 3.87 .01	2.51 3.01 3.70 .01	2.42 2.84 3.56 .01	2.36 2.78 3.46 .01	2.30 2.72 3.37 .01	2.25 2.64 3.17 .01	.05 .025 .01 20
24	4.26 5.72 7.62 .01	3.40 4.32 5.61 .01	3.01 3.72 4.72 .01	2.78 3.38 4.22 .01	2.62 3.15 3.90 .01	2.51 2.99 3.67 .01	2.42 2.87 3.50 .01	2.36 2.78 3.36 .01	2.30 2.70 3.26 .01	2.25 2.64 3.17 .01	2.22 2.59 3.09 .01	.05 .025 .01 24
30	4.17 5.57 7.56 .01	3.32 4.18 5.39 .01	2.92 3.59 4.51 .01	2.69 3.26 4.02 .01	2.53 3.03 3.70 .01	2.42 2.87 3.47 .01	2.33 2.75 3.30 .01	2.27 2.65 3.17 .01	2.21 2.57 3.07 .01	2.16 2.51 2.98 .01	2.13 2.46 2.90 .01	.05 .025 .01 30
40	4.08 5.42 7.31 .01	3.23 4.05 5.18 .01	2.84 3.46 4.31 .01	2.61 3.13 3.83 .01	2.45 2.90 3.51 .01	2.34 2.76 3.29 .01	2.25 2.62 3.12 .01	2.18 2.53 2.99 .01	2.12 2.45 2.89 .01	2.08 2.39 2.80 .01	2.04 2.33 2.73 .01	.05 .025 .01 40
60	4.00 5.29 7.08 .01	3.15 3.93 4.98 .01	2.76 3.34 4.13 .01	2.53 3.01 3.65 .01	2.37 2.79 3.34 .01	2.25 2.63 3.12 .01	2.17 2.51 2.93 .01	2.10 2.42 2.82 .01	2.04 2.33 2.72 .01	1.99 2.27 2.63 .01	1.95 2.22 2.56 .01	.05 .025 .01 60
120	3.92 5.15 6.85 .01	3.07 3.80 4.79 .01	2.68 3.23 3.95 .01	2.45 2.89 3.48 .01	2.29 2.67 3.17 .01	2.17 2.52 2.96 .01	2.09 2.39 2.79 .01	2.02 2.30 2.66 .01	1.96 2.22 2.56 .01	1.91 2.16 2.47 .01	1.87 2.10 2.40 .01	.05 .025 .01 120
$\infty$	3.84 5.02 6.63 .01	3.00 3.69 4.61 .01	2.60 3.11 3.78 .01	2.37 2.79 3.32 .01	2.21 2.57 3.02 .01	2.10 2.41 2.80 .01	2.01 2.29 2.64 .01	1.94 2.19 2.51 .01	1.88 2.11 2.41 .01	1.83 2.05 2.32 .01	1.79 1.99 2.25 .01	.05 .025 .01 $\infty$

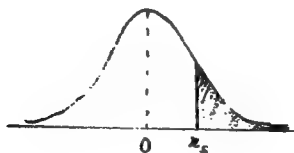
تابع الجدول (٩)  
القيم المحرجة لتوزيع ف

$\alpha$	12	15	20	24	30	40	60	120	$\infty$	$\alpha$
11 .05 .025 .01	2.79 3.43 4.40	2.72 3.33 4.23	2.65 3.23 4.10	2.61 3.17 4.02	2.57 3.12 3.94	2.53 3.06 3.86	2.49 3.00 3.78	2.45 2.94 3.69	2.40 2.88 3.60	.05 11 .025 .01
12 .05 .025 .01	2.69 3.28 4.16	2.62 3.18 4.01	2.54 3.07 3.86	2.51 3.02 3.78	2.47 2.96 3.70	2.43 2.91 3.62	2.38 2.85 3.54	2.34 2.79 3.45	2.30 2.72 3.36	.05 12 .025 .01
15 .05 .025 .01	2.48 2.96 3.67	2.40 2.86 3.52	2.33 2.76 3.37	2.39 2.70 3.29	2.25 2.64 3.21	2.20 2.59 3.13	2.16 2.52 3.05	2.11 2.46 2.96	2.07 2.40 2.87	.05 15 .025 .01
20 .05 .025 .01	2.28 2.68 3.23	2.20 2.57 3.09	2.12 2.46 2.94	2.08 2.41 2.86	2.04 2.35 2.78	1.99 2.29 2.69	1.95 2.22 2.61	1.90 2.16 2.52	1.84 2.09 2.42	.05 20 .025 .01
24 .05 .025 .01	2.18 2.54 3.03	2.11 2.44 2.89	2.03 2.33 2.74	1.98 2.27 2.66	1.94 2.21 2.58	1.89 2.15 2.49	1.84 2.08 2.40	1.79 2.01 2.31	1.73 1.94 2.21	.05 24 .025 .01
30 .05 .025 .01	2.09 2.41 2.84	2.01 2.31 2.70	1.93 2.20 2.55	1.89 2.14 2.47	1.84 2.07 2.39	1.79 2.01 2.30	1.74 1.94 2.21	1.68 1.87 2.11	1.62 1.79 2.01	.05 30 .025 .01
40 .05 .025 .01	2.04 2.29 2.66	1.92 2.18 2.52	1.84 2.07 2.37	1.79 2.01 2.29	1.74 1.94 2.20	1.69 1.88 2.11	1.64 1.80 2.02	1.58 1.72 1.92	1.51 1.64 1.80	.05 40 .025 .01
60 .05 .025 .01	1.92 2.17 2.50	1.84 2.06 2.35	1.75 1.94 2.20	1.70 1.88 2.12	1.65 1.82 2.03	1.59 1.74 1.94	1.53 1.67 1.84	1.47 1.58 1.73	1.39 1.48 1.60	.05 60 .025 .01
120 .05 .025 .01	1.83 2.05 2.34	1.75 1.95 2.19	1.66 1.82 2.03	1.61 1.76 1.95	1.55 1.69 1.86	1.50 1.61 1.76	1.43 1.53 1.66	1.35 1.43 1.53	1.25 1.31 1.38	.05 120 .025 .01
$\infty$ .05 .025 .01	1.75 1.94 2.18	1.67 1.83 2.04	1.57 1.71 1.88	1.52 1.64 1.79	1.46 1.57 1.70	1.39 1.48 1.59	1.32 1.39 1.47	1.22 1.27 1.32	1.00 1.00 1.00	.05 $\infty$ .025 .01



الجدول (١٠)  
القيم الحرجة لمعامل ارتباط الرتب (سبيرمان)

$n$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
5	0.900	—	—	—
6	0.829	0.886	0.943	—
7	0.714	0.786	0.893	—
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478



# الجدول (١١)

التحويل  $E = \frac{1}{4} \log \frac{1 + \mu}{1 - \mu}$  لعامل الارتباط

	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080
1	100	.110	.121	.131	.141	.151	.161	.172	.182
2	203	.213	.224	.234	.245	.255	.266	.277	.288
3	.310	.321	.332	.343	.354	.365	.377	.388	.400
4	.424	.436	.448	.460	.472	.485	.497	.510	.523
5	.549	.563	.576	.590	.604	.618	.633	.648	.662
6	.693	.709	.725	.741	.758	.775	.793	.811	.829
7	.867	.887	.908	.929	.950	.973	.996	1.020	1.045
8	1.099	1.127	1.157	1.188	1.221	1.256	1.293	1.333	1.376
	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008
90	1.472	1.478	1.483	1.488	1.494	1.499	1.505	1.510	1.516
91	1.528	1.533	1.539	1.545	1.551	1.557	1.564	1.570	1.576
92	1.589	1.596	1.602	1.609	1.616	1.623	1.630	1.637	1.644
93	1.658	1.666	1.673	1.681	1.689	1.697	1.705	1.713	1.721
94	1.738	1.747	1.756	1.764	1.774	1.783	1.792	1.802	1.812
95	1.832	1.842	1.853	1.863	1.874	1.886	1.897	1.909	1.921
96	1.946	1.959	1.972	1.986	2.000	2.014	2.029	2.044	2.060
97	2.092	2.109	2.127	2.146	2.165	2.185	2.205	2.227	2.249
98	2.298	2.323	2.351	2.380	2.410	2.443	2.477	2.515	2.555
99	2.646	2.700	2.759	2.826	2.903	2.994	3.106	3.250	3.453

المجدول (١٢)  
قيم معامل الارتباط  $r$  بدلالة  $n$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
1	.100	.110	.119	.129	.139	.149	.159	.168	.178	.187
2	.197	.207	.216	.226	.236	.245	.254	.264	.273	.282
3	.291	.300	.310	.319	.327	.336	.345	.354	.363	.371
4	.380	.389	.397	.405	.414	.422	.430	.438	.446	.454
5	.462	.470	.478	.485	.493	.500	.508	.515	.523	.530
6	.537	.544	.551	.558	.565	.572	.578	.585	.592	.598
7	.604	.611	.617	.623	.629	.635	.641	.647	.653	.658
8	.664	.670	.675	.680	.686	.691	.696	.701	.706	.711
9	.716	.721	.726	.731	.735	.740	.744	.749	.753	.757
1.0	.762	.766	.770	.774	.778	.782	.786	.790	.793	.797
1.1	.800	.804	.808	.811	.814	.818	.821	.824	.828	.831
1.2	.834	.837	.840	.843	.846	.848	.851	.854	.856	.859
1.3	.862	.864	.867	.869	.872	.874	.876	.879	.881	.883
1.4	.885	.888	.890	.892	.894	.896	.898	.900	.902	.903
1.5	.905	.907	.909	.910	.912	.914	.915	.917	.919	.920
1.6	.922	.923	.925	.926	.928	.929	.930	.932	.934	.934
1.7	.935	.937	.938	.939	.940	.941	.942	.944	.945	.946
1.8	.947	.948	.949	.950	.951	.952	.953	.954	.954	.955
1.9	.956	.957	.958	.959	.960	.960	.961	.962	.963	.963
2.0	.964	.965	.965	.966	.967	.967	.968	.969	.969	.970
2.1	.970	.971	.972	.972	.973	.973	.974	.974	.975	.975
2.2	.976	.976	.977	.977	.978	.978	.978	.979	.979	.980
2.3	.980	.980	.981	.981	.982	.982	.982	.983	.983	.983
2.4	.984	.984	.984	.985	.985	.985	.986	.986	.986	.986
2.5	.987	.987	.987	.987	.988	.988	.988	.988	.989	.989
2.6	.989	.989	.989	.990	.990	.990	.990	.990	.991	.991
2.7	.991	.991	.991	.992	.992	.992	.992	.992	.992	.992
2.8	.993	.993	.993	.993	.993	.993	.993	.994	.994	.994
2.9	.994	.994	.994	.994	.994	.995	.995	.995	.995	.995

الجدول (١٣)  
الاحتمالات المتجمعة في اختبار التلاحقات

$(n_1, n_2)$	$\alpha$								
	2	3	4	5	6	7	8	9	10
(2, 3)	0.200	0.500	0.900	1.000					
(2, 4)	0.133	0.400	0.800	1.000					
(2, 5)	0.095	0.333	0.714	1.000					
(2, 6)	0.071	0.286	0.643	1.000					
(2, 7)	0.056	0.250	0.583	1.000					
(2, 8)	0.044	0.222	0.533	1.000					
(2, 9)	0.036	0.200	0.491	1.000					
(2, 10)	0.030	0.182	0.455	1.000					
(3, 3)	0.100	0.300	0.700	0.900	1.000				
(3, 4)	0.057	0.200	0.543	0.800	0.971	1.000			
(3, 5)	0.036	0.143	0.429	0.714	0.929	1.000			
(3, 6)	0.024	0.107	0.345	0.643	0.881	1.000			
(3, 7)	0.017	0.083	0.283	0.583	0.833	1.000			
(3, 8)	0.012	0.067	0.236	0.533	0.788	1.000			
(3, 9)	0.009	0.055	0.200	0.491	0.745	1.000			
(3, 10)	0.007	0.045	0.171	0.455	0.706	1.000			
(4, 4)	0.029	0.114	0.371	0.629	0.886	0.971	1.000		
(4, 5)	0.016	0.071	0.262	0.500	0.786	0.929	0.992	1.000	
(4, 6)	0.010	0.048	0.190	0.405	0.690	0.881	0.976	1.000	
(4, 7)	0.006	0.033	0.142	0.333	0.606	0.833	0.954	1.000	
(4, 8)	0.004	0.024	0.109	0.279	0.533	0.788	0.929	1.000	
(4, 9)	0.003	0.018	0.085	0.236	0.471	0.745	0.902	1.000	
(4, 10)	0.002	0.014	0.068	0.203	0.419	0.706	0.874	1.000	
(5, 5)	0.008	0.040	0.167	0.357	0.643	0.833	0.960	0.992	1.000
(5, 6)	0.004	0.024	0.110	0.262	0.522	0.738	0.911	0.976	0.998
(5, 7)	0.003	0.015	0.076	0.197	0.424	0.652	0.854	0.955	0.992
(5, 8)	0.002	0.010	0.054	0.152	0.347	0.576	0.793	0.929	0.984
(5, 9)	0.001	0.007	0.039	0.119	0.287	0.510	0.734	0.902	0.972
(5, 10)	0.001	0.005	0.029	0.095	0.239	0.455	0.678	0.874	0.958
(6, 6)	0.002	0.013	0.067	0.175	0.392	0.608	0.825	0.933	0.987
(6, 7)	0.001	0.008	0.043	0.121	0.296	0.500	0.733	0.879	0.966
(6, 8)	0.001	0.005	0.028	0.086	0.226	0.413	0.646	0.821	0.937
(6, 9)	0.000	0.003	0.019	0.063	0.175	0.343	0.566	0.762	0.902
(6, 10)	0.000	0.002	0.013	0.047	0.137	0.288	0.497	0.706	0.864
(7, 7)	0.001	0.004	0.025	0.078	0.209	0.383	0.617	0.791	0.922
(7, 8)	0.000	0.002	0.015	0.051	0.149	0.296	0.514	0.704	0.867
(7, 9)	0.000	0.001	0.010	0.035	0.108	0.231	0.427	0.622	0.806
(7, 10)	0.000	0.001	0.006	0.024	0.080	0.182	0.355	0.549	0.743
(8, 8)	0.000	0.001	0.009	0.032	0.100	0.214	0.405	0.595	0.786
(8, 9)	0.000	0.001	0.005	0.020	0.069	0.157	0.319	0.500	0.702
(8, 10)	0.000	0.000	0.003	0.013	0.048	0.117	0.251	0.419	0.621
(9, 9)	0.000	0.000	0.003	0.012	0.044	0.109	0.238	0.399	0.601
(9, 10)	0.000	0.000	0.002	0.008	0.029	0.077	0.179	0.319	0.510
(10, 10)	0.000	0.000	0.001	0.004	0.019	0.051	0.128	0.242	0.414

تابع الجدول (١٣)  
الاحتمالات المتجمعة في اختبار التلاحقات

$(n_1, n_2)$	$\alpha$									
	11	12	13	14	15	16	17	18	19	20
(2, 3)										
(2, 4)										
(2, 5)										
(2, 6)										
(2, 7)										
(2, 8)										
(2, 9)										
(2, 10)										
(3, 3)										
(3, 4)										
(3, 5)										
(3, 6)										
(3, 7)										
(3, 8)										
(3, 9)										
(3, 10)										
(4, 4)										
(4, 5)										
(4, 6)										
(4, 7)										
(4, 8)										
(4, 9)										
(4, 10)										
(5, 5)										
(5, 6)	1.000									
(5, 7)	1.000									
(5, 8)	1.000									
(5, 9)	1.000									
(5, 10)	1.000									
(6, 6)	0.998	1.000								
(6, 7)	0.992	0.999	1.000							
(6, 8)	0.984	0.998	1.000							
(6, 9)	0.972	0.994	1.000							
(6, 10)	0.958	0.990	1.000							
(7, 7)	0.975	0.996	0.999	1.000						
(7, 8)	0.949	0.988	0.998	1.000	1.000					
(7, 9)	0.916	0.975	0.994	0.999	1.000					
(7, 10)	0.879	0.957	0.990	0.998	1.000					
(8, 8)	0.900	0.968	0.991	0.999	1.000	1.000				
(8, 9)	0.843	0.939	0.980	0.996	0.999	1.000	1.000			
(8, 10)	0.782	0.903	0.964	0.990	0.998	1.000	1.000			
(9, 9)	0.762	0.891	0.956	0.988	0.997	1.000	1.000	1.000		
(9, 10)	0.681	0.834	0.923	0.974	0.992	0.999	1.000	1.000	1.000	
(10, 10)	0.586	0.758	0.872	0.949	0.981	0.996	0.999	1.000	1.000	1.000

الجدول (١٤)

القيم الحرجة لاختبار ويلكوكسون للمقارنات التزاوجية

$n$	One-sided $\alpha = 0.01$ Two-sided $\alpha = 0.02$	One-sided $\alpha = 0.025$ Two-sided $\alpha = 0.05$	One-sided $\alpha = 0.05$ Two-sided $\alpha = 0.10$
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11
11	7	11	14
12	10	14	17
13	13	17	21
14	16	21	26
15	20	25	30
16	24	30	36
17	28	35	41
18	33	40	47
19	38	46	54
20	43	52	60
21	49	59	68
22	56	66	75
23	62	73	83
24	69	81	92
25	77	90	101
26	85	98	110
27	93	107	120
28	102	117	130
29	111	127	141
30	120	137	152

† Reproduced from F. Wilcoxon and R. A. Wilcoxon *Some Rapid Approximate Statistical Procedures*, American Cyanamid Company, Pearl River, N.Y., 1964 by permission of the American Cyanamid Company

# الجدول (١٥)

## الاحتمالات ل (س ≤ س) في اختيار الاتجاه

For  $n = 3$ ,  $F(2) = 1 - 0.167 = 0.833$ .

For  $n = 4$ ,  $F(3) = 1 - 0.375 = 0.625$ ,  $F(4) = 1 - 0.167 = 0.833$ , etc.

$x$	$n=3$	$x$	$n=4$	$x$	$n=5$	$x$	$n=6$	$x$	$n=7$	$x$	$n=8$	$x$	$n=9$	$x$	$n=10$	$x$	$n=11$
0	0.167	0	0.042	0	0.008	0	0.001	1	0.001	2	0.001	4	0.001	6	0.001	8	0.001
1	500	1	167	2	117	2	028	3	015	4	007	5	003	7	002	9	002
		2	375	3	242	3	068	4	035	5	016	7	012	9	008	11	005
				4	408	4	136	5	068	6	031	7	022	10	014	12	008
						5	235	6	119	7	054	9	038	11	023	13	013
						6	360	7	191	8	089	10	060	12	036	14	020
						7	500	8	281	9	138	11	090	13	054	15	030
								9	386	10	199	12	130	14	078	16	043
								10	500	11	274	13	179	15	108	17	060
										12	360	14	238	16	146	18	082
										13	452	15	306	17	190	19	109
												16	381	18	242	20	141
												17	460	19	300	21	179
														20	364	22	223
														21	431	23	271
														22	500	24	324
																25	381
																26	440
																27	500

$x$	$n=19$	$x$	$n=18$	$x$	$n=17$	$x$	$n=16$	$x$	$n=15$	$x$	$n=14$	$x$	$n=13$	$x$	$n=12$
50	0.001	43	0.001	38	0.001	32	0.001	2	0.001	23	0.001	18	0.001	14	0.001
51	0.002	44	0.002	39	0.002	33	0.002	28	0.002	24	0.002	19	0.002	15	0.001
52	0.002	45	0.002	40	0.003	34	0.002	29	0.002	25	0.003	20	0.002	16	0.002
53	0.003	46	0.003	41	0.003	35	0.003	30	0.003	26	0.004	21	0.003	17	0.003
54	0.004	47	0.003	42	0.004	36	0.004	31	0.004	27	0.006	22	0.005	18	0.005
55	0.005	48	0.004	43	0.005	37	0.005	32	0.006	28	0.008	23	0.007	19	0.007
56	0.006	49	0.005	44	0.007	38	0.007	33	0.008	29	0.010	24	0.010	20	0.011
57	0.007	50	0.006	45	0.009	39	0.009	34	0.010	30	0.012	25	0.013	21	0.015
58	0.008	51	0.008	46	0.011	40	0.011	31	0.011	31	0.014	26	0.018	22	0.022
59	0.010	52	0.010	47	0.013	41	0.014	32	0.012	32	0.016	27	0.024	23	0.028
60	0.012	53	0.012	48	0.016	42	0.017	33	0.013	33	0.020	28	0.031	24	0.033
61	0.014	54	0.014	49	0.020	43	0.021	34	0.016	34	0.025	29	0.040	25	0.040
62	0.017	55	0.017	50	0.024	44	0.026	35	0.016	35	0.030	30	0.051	26	0.050
63	0.020	56	0.021	51	0.029	45	0.032	36	0.016	36	0.035	31	0.063	27	0.058
64	0.023	57	0.025	52	0.034	46	0.038	37	0.021	37	0.040	32	0.079	28	0.066
65	0.027	58	0.029	53	0.041	47	0.046	38	0.026	38	0.045	33	0.096	29	0.076
66	0.032	59	0.034	54	0.048	48	0.054	39	0.032	39	0.050	34	0.117	30	0.088
67	0.037	60	0.040	55	0.056	49	0.064	40	0.039	40	0.055	35	0.140	31	0.103
68	0.043	61	0.047	56	0.066	50	0.076	41	0.048	41	0.060	36	0.165	32	0.125
69	0.049	62	0.054	57	0.076	51	0.088	42	0.058	42	0.065	37	0.194	33	0.155
70	0.056	63	0.062	58	0.088	52	0.102	43	0.070	43	0.070	38	0.225	34	0.190
71	0.064	64	0.072	59	0.100	53	0.118	44	0.083	44	0.075	39	0.259	35	0.230
72	0.073	65	0.082	60	0.115	54	0.135	45	0.097	45	0.080	40	0.295	36	0.273
73	0.082	66	0.093	61	0.130	55	0.154	46	0.114	46	0.090	41	0.334	37	0.319
74	0.093	67	0.105	62	0.147	56	0.174	47	0.133	47	0.100	42	0.374	38	0.369
75	0.104	68	0.119	63	0.165	57	0.196	48	0.153	48	0.110	43	0.415	39	0.429
76	0.117	69	0.133	64	0.184	58	0.220	49	0.175	49	0.120	44	0.457	40	0.476
77	0.130	70	0.149	65	0.205	59	0.245	50	0.199	50	0.130	45	0.500		
78	0.144	71	0.166	66	0.227	60	0.271	51	0.225	51	0.140				
79	0.159	72	0.184	67	0.250	61	0.299	52	0.253	52	0.150				
80	0.176	73	0.203	68	0.275	62	0.328	53	0.282	53	0.160				
81	0.193	74	0.223	69	0.300	63	0.358	54	0.313	54	0.170				
82	0.211	75	0.245	70	0.327	64	0.388	55	0.345	55	0.180				
83	0.230	76	0.267	71	0.354	65	0.420	56	0.378	56	0.190				
84	0.250	77	0.290	72	0.383	66	0.452	57	0.412	57	0.200				
85	0.271	78	0.314	73	0.411	67	0.484	58	0.447	58	0.210				
86	0.293	79	0.339	74	0.441			59	0.482	59	0.220				
87	0.315	80	0.363	75	0.470										
88	0.339	81	0.391												
89	0.362	82	0.418												
90	0.387	83	0.445												
91	0.411	84	0.473												
92	0.436	85	0.500												
93	0.462														
94	0.487														





## References

- Bhattacharyya, G. and Johnson R.:** Statistical Concepts and Methods, Wiley, 1977.
- Bishop, O. N.:** Statistics for Biology, Longman, 1971.
- Cochran, W. G.:** Sampling Techniques, Wiley, 1977.
- Cochran, W. G. and Cox, G. M.:** Experimental Designs, Wiley, 1957.
- Colquhoun, D.:** Lectures on Biostatistics, Clarendon Press, Oxford, 1971.
- Freund, J. E.:** Modern Elementary Statistics, Pentice - Hall, 1987.
- Guust, R. E. and Mason, R.L.:** Regression Analysis and Application.
- Hays, W. L.:** Statistics for Social Sciences, Holt, Rinehart and Winston, 1973.
- Hill, A. B.:** A Short Textbook of Medical Statistics, Lancet, 1984.
- Kreyszig, E.:** Introductory Mathematical Statistics, Wiley, 1970.
- Krumben, W. C. and Graybill, F. A.:** Statistical Models in Geology, McGraw-Hill, 1965.
- Milton, J. S. and Others:** Introduction to Statistics, Heath and Company, 1986.
- Nalimov, V. V.:** The Application of Mathematical Statistics to Chemical Analysis.
- Skane, D. H.:** Elementary Statistics, Addison-Wesley, 1985.
- Sokal, R. R. and Rohlf, F. J.:** Introduction to Biostatistics, Freeman, 1973.
- Snedecor, G. W. and Cochran, W. G.:** Statistical Methods, Iowa State Univ. Press, 1980.
- Sykes, M. N. and Vickers, M. D.:** Principles of Clinical Measurement, McGraw-Hill, 1982.
- Walpole, R. and Myers, R.:** Probability and Statistics for Engineers and Scientists, Macmillan, 1978.
- Wardlaw, A.C:** Practical Statistics for Experimental Biologists.
- Winer, B. J.:** Statistical Principles in Experimental Designs, McGraw-Hill, 1971.
- Zuwaylif, F. H.:** Applied Business Statistics, Addison-Wesley, 1984.







يعرض هذا الكتاب شيئاً مما يسمى بالإحصاء التطبيقي ، وهو يتناول جل المفاهيم والطرق والنماذج الإحصائية التي يحتاج إليها الطلاب والباحثون التطبيقيون في مختلف ميادين البحث العلمي . وتستند معالجة الموضوعات المقدمة على أمثلة توضيحية تيسر للقارئ تقبل ما يعرض من مادة وتضيء له طريق استخدامها في التطبيقات العملية . وتدعم هذه المعالجة بتأريخ صممت لتستثير فكر القارئ وتعاونه على ربط النقاط الأساسية فيما يقرأ ، كما تمنحه الفرصة لتقويم ما استوعبه من المادة وتحسين هذا الاستيعاب عن طريق التغذية المرتجعة التي توفرها الأجوبة الشاملة المعطاه بالملحق (١) لهذا الكتاب .

والناشر إذ يفخر بتقديم هذا المرجع القيم يرجو أن يكون فيه إضافة جوهريّة للمكتبة العربية .



Bibliotheca Alexandrina



0255474